

## **CHAPTER I**

### **Overview**

#### **1.0 Background**

Speech technologies continue to advance and improve in capability. This has allowed speech-enabled interfaces to be applied in many different consumer and business applications throughout the world. Applications range from voice-controlled toys to nationwide use of voice dialing in wireless networks. This report will focus on the major opportunities that telephony applications represent for speech technology markets, both in public telephony networks and in improving telecommunications services for enterprises, especially their call centers.

In the public networks, speech recognition technologies have been implemented to create a number of new services that make the telephone a more dynamic medium. Speech recognition affects four major telephony areas: voice dialing, voice portals, unified messaging and in-car communications, called telematics.

Most of the attention is on voice-activated information content, created as services called voice portals. Although the phrase, voice portals, is used in various ways, essentially they are information databases that can be navigated by voice commands. In the current version, users of voice portals dial a toll-free phone number on their wireline or wireless phones and use spoken commands to listen to various kinds of information, including current traffic reports, stock quotes, flight information and up-to-date news stories.

However, the basic use of speech technology in simply making calls is expected to become widespread. It seems a natural progression for telephony services to add voice commands to either basic dialing or managing calling features or enhanced services. Voice-activated commands simply function as a replacement for keypad punching. Speech-enabled interfaces are especially important in the wireless world, since mobile telephony services are geared toward the person who is on the move. While on the move, working on the small keypads that are found on today's tiny and lightweight mobile phones is not an easy task. Plus, current regulatory sentiment seeks restrictions on the unlimited use of mobile devices in vehicles. This means voice dialing and voice interfaces will most likely become a necessity in that major venue. Already New York State has passed a law that became effective November 1, 2001 that forbids motorists from using a handheld cell phone while driving. Hands-free talking or headsets are permitted.

Speech technology not only helps with making wireless voice calls, but is being deployed as an interface for wireless data applications as well. Mobile phones and other devices are being used for access to email, the Web and other wireless information services. This is important since wireless data use is growing. Wireless carriers are using enhanced services, such as information services and wireless Web access, as tools to differentiate themselves in the marketplace. But the information services that people on-the-move access, need to be easy to use. It is widely believed that speech recognition will become the key way for carriers to provide greater ease-of-use to their wireless data subscribers.

Enterprises have found a number of valuable uses for speech technology. Speech technologies can help corporations improve their internal telephone services and mobile systems. Speech technologies may be used on a company's general phone number for speech-enabled call routing, auto attendant, directory assistance, and virtual personal assistant. These general applications have made use of interactive voice response (IVR) technologies in the past, but are now being improved with the addition of speech recognition. This has allowed firms to expand automation to include voice-enabled

access to email and voice mail, along with voice-activated dialing. These applications have also been applied in the areas of field force automation (FFA) and customer relationship management (CRM).

One of the most widely applied and proven areas of speech technology in enterprises is its application in call centers (contact centers). Throughout the world, speech technologies in call centers are used to cut costs, improve customer service, and improve sales. There is a growing amount of evidence for successful ROIs in application of speech technologies to call centers. Nuance chronicles the success at United Airlines' toll-free center, for example, where up to 70 percent of flight inquiries are now handled by its automated, speech-enabled system, saving them millions of dollars in personnel and overhead costs. The first types of firms to adopt the new technology have been in the travel and brokerage industries, but now firms in banking, insurance, and utilities have been adopting it. Government offices and agencies are also taking advantage of speech technologies.

### **1.1 Explanation of Speech Technologies**

To understand how speech technologies will impact telephony markets, it is important to understand the vast difference between implementations of interactive voice response (IVR) or the older speech software and the newer speech recognition systems. IVR has been around for a number of years and is already widely applied in call centers and directory assistance applications. It allows callers to enter database queries using a touchtone pad. Users navigate the system by listening to software-generated voice prompts and then making choices using the keypad. The information they hear is spoken back to them using speech software, but the user commands are done by keypad and not by voice. Although this interface is easy to use on a regular phone keypad, some people have grown frustrated with the task, especially when there are long menus and extended hierarchical paths. The task is even more frustrating for users of handheld mobile phones and other portable devices, where it can be quite

irksome and even difficult to use. IVR had its day. But speech vendors are now able to bring to bear advanced speech technology that enables the creation of applications that present end-to-end automated voice interactions without the use of touchtone.

For spoken commands to work easily and effectively in everyday usage, speech technology had to overcome some key technology hurdles. All the speech labs around the world have had to face these same hurdles, since they basically used the same overall approach in creating speech recognition. Speech recognition works as follows. First, the speech recognition system breaks up incoming voice signals into sound wave patterns. Next a software algorithm is used to correctly match the sound waves against a stored database of known speech waves. This matching algorithm is very important, since it acts as the translator. It translates the incoming spoken words into words that an automated system can recognize.

Equally important are the stored wave patterns. Speech recognition vendors have developed huge databases of speech waves that comprise patterns for thousands of words, in all forms of diction and even foreign accents or foreign languages. This huge database of stored wave patterns is used as the database from which the speech recognition engine pulls out the recognized word. The spoken word coming from a telephone caller is matched against the stored ones in order to find the closest match. Recognizing a word such as a name, or really the way someone says a name, has to work in a highly predictable way.

In the first years of speech research, if someone "taught" the software to recognize their individual voice, basic voice commands worked fairly reliably. The system could be taught to recognize someone's words by repetition. By repeating some basic word patterns a number of times into the speech database, the individual could actually teach the machine to recognize his or her voice. The aim was accuracy. Accuracy is key, because people get frustrated if they have to repeat a word more than once, or if the service does not recognize the word at all.

Some of the early versions of speaker-dependent software were applied in voice dialing chip processors in telephone handsets. Limited vocabulary sets of speech

recognition were also incorporated into spoken routing commands applied to call center service applications. Carriers have offered services based on speech recognition, such as voice dialing and voice-activated voice mail, for several years. However, these applications only required the speech recognition software to recognize nine digits for telephone numbers or a limited set of commands related to accessing voice messages. Only recently have improvements in the vocabularies and algorithms of speech technology allowed for a wider set of commands and functions.

And only recently has the reliability improved to allow it to work with any voice, not just the subscriber who has trained the system to his or her own voice. This is known as speaker independent speech software. In large measure this has occurred because of two simultaneous events: advancements in speech software and the tremendous increase in processing power. As one benchmark, BeVocal's Mikael Berner said that recent advancements in filtering and speech software have improved accuracy rates to 90 percent or better. Some vendors report that they have successfully applied speech in calls centers for 95 or 96 percent accuracy. Secondly, memory and processing speeds have been doubling every year and the costs have been reduced by 50 percent per year. This means that vendors can create large enough and fast enough databases to accomplish all that is necessary in storing voice wave patterns and running the recognition algorithms.

### **1.1.1 Definition**

A number of phrases are used to describe the capabilities of speech technology: natural language, voice recognition, speech recognition, voice activation and others. The core technology that is igniting markets is speech recognition -- the robust ability of speech servers in the network to recognize spoken words. Speech technologies also include two other important capabilities: text-to-speech (stored text is converted to spoken words or vice versa) and voice authentication (someone's unique voice is verified using biometric technology). Carriers and their partners are beginning to

incorporate some of these new capabilities into their networks and services. These capabilities go beyond simple phone dialing and basic call command sequences. These capabilities allow for directed dialogues to be implemented into networks to allow for information retrieval. Enterprises can also buy solutions for private networks and geared toward the communications needs of employees and those that call into their company, such as customers and suppliers.

In the basic version of retrieval, a speech recognition system allows callers to speak to the system by answering questions. This can be done in a simple way through directed dialogues. In directed dialogues, the system directs the caller through a series of questions to arrive at the needed information. The system only requires the user to respond with a single word. For example, the system might ask. "Do you want to access the phone directory, yes or no?" The user simply says either yes or no. They then move on to the next step.

At the next level, the more sophisticated systems allow a more natural approach. Users can speak to the system in sentences. For example, instead of saying "checking account", callers can say, "I'd like my checking account balance." The highest level of speech technology is the use of more complex sentences that permit multiple operations to occur at once. A caller can say "I want to transfer \$500 dollars from my savings account to my checking account." This is one of the major innovations of the current technology. Users can direct the system to do several activities at once.

New systems also allow for what is called barge-in. This allows callers to interrupt the system, while it is in the midst of prompting them. While the system is still suggesting choices, it can still pick up and understand a user's words that interrupt it. This allows the repeat user to speed up their interactions with the system.

Speaker verification or authentication is another capability of speech technology. This feature will identify a caller based on the characteristics of their unique voice patterns. Speaker verification software contains a program that can be trained to recognize an individual's voice. For applications such as portfolio management, this is a key feature. Voiceprints, like fingerprints, are extremely accurate and often companies

do not need any further proof of identification. Speaker verification can be used for security purposes to verify the identity of someone requesting sensitive or confidential information.

### **1.1.2 Adoption of Speech Technologies**

Speech technologies have come of age because of improved software and computing power. Companies are seizing new opportunities related to speech technology because the core technologies have improved and standard voice-interface protocols have been established. After many years of research, speech recognition software from companies such as Lucent, IBM, Nuance and SpeechWorks, has become robust enough to be used in mass audience applications. Current versions of speech recognition software are now able to understand a wide range of accents, diction and speech patterns.

In the past, speech recognition software was useful, but it had to be trained by each individual user before it was accurate. And even with that, it was not all that reliable in normal working situations. The new speech software is now highly accurate. It works in most situations -- working with most voices and filtering out background noises. The other types of speech software have also matured including text-to-speech, speech-to-text and speaker verification software. Natural language software, which allows users to ask questions in complete sentences, is just emerging and initial applications show promises that will make speech-enabled applications even easier to use.

The other critical factor in the promising potential for speech technology is the fact that standards are being developed. Consortiums are in place to foster new voice-related computer protocols. A computer language related to voice interfaces, called VoiceXML, makes it easy to write new voice-activated services in computer language. As an accepted standard, it makes creating voice portals straightforward.

Technology firms and carriers alike have invested in VoiceXML. Just as HTML is used to create Web pages or WAP is used to create Web views on mobile devices, VoiceXML will allow voice access to services. Using VoiceXML language, messages or data will pass through a speech server with VoiceXML interpreters so that the data is voice-activated. Users can navigate by voice and "read" the information through text-to-speech converters.

With VoiceXML, the human voice becomes a substitute for a computer mouse or a telephone keypad. When a user speaks into their wireline or wireless phone, they can call up content, which is specially designed for telephone access. The content may come in the form of sound clips, numbers, music, and of course, spoken text. Voice recognition technology is also being used for hands-free dialing and more efficient message and information retrieval on wireless handsets and in telematics units in automobiles.

## **1.2 Speech Technologies Used in Telephony**

Applying speech to telephone networks, call centers or enterprises leverages current investments in phone systems and personal while at the same time invoking higher levels of customer satisfaction and creating new services. Companies are also leveraging the use of their legacy databases, either directories or customer record databases, using speech technologies. Customer service operations often need to access a data source for retrieving information on customers or conducting transactions. In many cases, the same database or Web site data source used by customer service reps can be used for both screen access and speech access solutions.

### **1.2.1 Speech in the Network**

Callers can be connected using either the landline or wireless networks, while enjoying the benefits of voice activation. The basic architecture for using speech in the



network is seen in Figure 1.1. It shows the caller being routed through the network and connected via a VoiceXML Gateway (sometimes referred to as speech gateway or voice gateway). Their voice call can be enhanced with speech-enabled telephony applications to services such as voice dialing and voice command. In addition, they can be connected via a voice browser to various Web and information services powered by voice activation.

### **1.2.2 Types of Speech Technology**

Speech technology consists of several types of core software that enables application developers in the telephony arena to apply speech-enabled mechanisms to their environments. Sometimes these different types of software are used separately and sometimes they are used in combination. The core software is usually licensed from a speech technology firm and is used by a variety of players, including application developers, ASPs, system integrators and the carriers themselves, in the creation of the end service.

Some of the software providers bundle tools that developers need to have to create, deploy and tune their speech services. Speech companies may also include software building blocks or application sets. These modules are packaged or bundled sets of common speech interactions. Modules have been created for simple exchanges such as yes/no sequences, request for credit cards, or phone number dialing. The use of prepackaged modules help developers save time and prevents the need to recreate code for common functions.

The following sections profile each type of core speech technology software.

#### **1.2.2.1 Automatic Speech Recognition (ASR)**

Automatic speech recognition (ASR) is software that enables a computer platform to understand what someone says. In telephony networks, callers on the phone can

speak digits, letters or words into a phone and the system captures their utterance. The system must then translate the utterance into something the system can interpret and understand -- and use in responding to the caller request.

In telephone networks, speech recognition platforms can be added into the network so that a caller from any location speaks and the speech recognition engine captures the utterance. The basic call flow is captured in Figure 1.2. It shows how a call is captured in the voice gateway and then processed to come up with the results.

Technically, the system digitizes the spoken voice and converts it into a spectral representation, which describes the way the caller's spoken words sound but broken into individual frequency segments or components. These segments of spectral representations are converted into phonemes, the basic sounds of language, such as the "m" in "man." These phonemes are used in the search and matching process that takes place in the speech database. The matching comes up with the best match for words. These matches are processed in order to come up with the understanding of the words that were spoken.

If you ask for only a few simple words, such as yes/no, stop, call, then the set of matches is small and the system can be highly accurate. But ASR continues to improve as the speech engines improve and the available sets of vocabularies grow. Using better statistical modeling techniques, accuracy of matching improves. Speech engines are also more robust in their ability to handle noise, accents and different languages.

Current interactive voice response (IVR) platforms can be improved with the addition of ASR. This is the major application of speech technology in enterprise call centers today. Callers into corporate directories or call centers find a greater utility with speech recognition. Instead of responding to voice prompts with touchtone (DTMF) commands, users can simply speak their responses. ASR has also been found useful in the processing of callers' requests in transaction-based applications. Speech commands are used to allow employees access to the information they need from an enterprise database or Web site.

### **1.2.2.2 Text-to-Speech (TTS)**

Text-to-speech software (TTS) is the other major speech technology finding application in telephony markets. TTS software is able to take any stored or input text and convert it to spoken sounds. Text of all kinds, from email messages to mainframe-generated text, is used as input and the output comes in audio form. TTS is implemented entirely in software modules and only standard audio capability is required.

Text-to-speech systems have various applications. These include reading electronic mail messages, generating spoken prompts in voice response systems, and as an interface to an order-verification system for salespeople in the field. Another primary use of TTS is to provide information content over the phone. This helps in cases where pre-recorded information is not practical. Examples include reading out addresses from a directory listing database or giving driving instructions from a geographic street-by-street navigation database. In the enterprise side, TTS is found in server and PBX applications, such as:

- Email reading;
- Fax reading;
- Directory and reverse directory listing;
- Voice portals;
- IVR; and
- Unified messaging.

Specialized enterprise uses can be found in applications where there is need for supervision or alerting systems by telephone and radio. For example, it has been applied in manufacturing plants where remote maintenance monitors detect a problem and then place a call to an emergency response center. The person answering the call gets the monitoring readout via TTS.

There are numerous speech software firms, VARs and resellers of speech-to-text software. Firms included in our detailed profiles are: AT&T Labs, BBN Technologies,

Elan, Fonix, Lernout & Hauspie, Lucent Speech Solutions, Nuance Communications, SpeechWorks, and Telisma. Other firms in TTS are: Babel Technologies, Conversay, E-Speech, Gold Systems, Paragon Voice Systems, Rhetorical Systems, Sail Labs Technology, Sensory Inc., Verascape, VoiceGenie Technologies, and West Corporation.

### **1.2.2.3 Speaker Verification**

Speaker verification software is sometimes referred to as speaker authentication or voice authentication software. Someone calling into a speaker verification platform can be verified as to their identity based on his or her voice patterns and characteristics. Speaker verification is a biometric technology. It is used in security applications and for voice portal deployments where personal identity must be identified up front.

The speaker verification process works by verifying the claimed identity of a registered speaker using his/her voice characteristics. The speaker needs to enroll before using the system. During enrollment, the speaker speaks a given set of utterances, which the system uses to build statistical models representing the speaker's voice. Each individual in the world has a unique voice pattern, called a voice print, which uniquely identifies him or her.

Speaker verification technology relies on unique biometric characteristics of a person's voice. This is an alternative to the use of IDs or passwords and is more efficient than requiring typed alphanumeric sequences. By comparing a caller's voiceprint with the reference voiceprint associated with them, speaker verification software can verify if the caller is authorized to access desired information. Speaker verification technology can actually provide two layers of authentication by requiring both the correct password to be used and by matching a user's voiceprint. Therefore, it is more secure than using personal identification numbers (PINs) or passwords alone.

Figure 1.3 shows the process for an integrated speaker verification call that includes both voice imprint and normal id/password security. In the example given here, the

user speaks their name (which is used to verify their voiceprint) and states their ID number for processing the customer ID database.

An example of speaker authentication is Nuance's Verifier 3.0. In Verifier 3.0, a caller registers his or her voice by answering a few questions. This process takes up to two minutes. Only a single enrollment is necessary for any number of access types (e.g. wireless, wireline, or VoIP). In order to access a secure area, to perform a transaction or to access personal information, the user simply provides his or her name or other information requested by the system and the person's voice is compared against the voiceprint on file. If there is a match, the caller is given access to the system or secure area. Nuance Verifier provides security even in the most challenging, noisy environments such as airports and other high-traffic areas. This is because a voiceprint is not simply an audio recording of a user's voice. Just as a fingerprint is a unique physical trait, so is a voiceprint. It is based on the physical characteristics of an individual's vocal tract and cannot be altered, stolen or forged.

The upgraded version of Nuance's speaker verification product was announced in the second half of 2001. Nuance Verifier 3.0 can eliminate security breaches for airlines, enterprises and government agencies through voice authentication. Compared to other biometrics, including fingerprinting, handprinting, iris, retina or facial scanning, voice authentication was proven as a superior security solution, showing a 99.9 percent accuracy rate in a recent research study commissioned by Communications Electronics Security Group (CESG) in the UK. Nuance Verifier performs voice authentication from any network access point with a microphone or ordinary telephone. No special equipment is necessary and the technology can handle a range of applications from facilities access to the protection of sensitive data.

Other speaker verification firms include: Appriss, BBN Technologies, Enterprise Verification Group, Fast-Talk Communications, GIMS Inter@active, Gold Systems, Lernout & Hauspie, Paragon Voice Systems, Persay, Sail Labs, Sensory, SpeechWorks, T-Netrix, Telisma, Veritel, and Vocentric.

#### 1.2.2.4 Natural Language Understanding (NLU)

Natural language understanding (NLU) (aka natural language processing) is built into a software engine so that it can process normal speech. NLU refers to the more natural speaking patterns of continuous, fluid speech without any pauses between each word. In NLU implementations, callers use complete sentences to invoke commands rather than respond one question at a time to various prompts. For example, speaking to a travel database, someone might say, "I am at the corner of 7<sup>th</sup> and Beeker in Newark, New Jersey and want to go to 213 Main Street in Hackensack, New Jersey." This would be captured, the words would be recognized using speech recognition, but then it would be interpreted by the natural language understanding software into the exact information needed to process the request.

Figure 1.4 shows the call flow for a system that has incorporated natural language processing. The system asks the user: "Please tell me the reason for your call today." The user then responds, "I have a question about my bill." The system must first use speech recognition to decipher the spoken words. This text output is then sent through the natural language engine, which detects which issue they want and routes the call, as normally, through the IVR system to give them the help they need.

The goal of natural language understanding is to design and build a system that will analyze, understand, and generate languages that humans use naturally, so that eventually one can address a computer or telephone as if one were addressing another person. This goal has not been fully achieved and will not be easy to reach. Recognizing words is one thing, but understanding spoken language another. Understanding language entails rather sophisticated mechanisms. It involves knowing what concepts a word or phrase stands for and knowing how to link those concepts together in a meaningful way. The researchers at Microsoft's speech lab use this example to explain the reason for this difficulty:

"As an English speaker you effortlessly understand a sentence like --  
Flying planes can be dangerous. Yet this sentence presents some

difficulties to a software program that lacks both your knowledge of the world and your experience with linguistic structures. Is the more plausible interpretation that the pilot is at risk, or that the danger is to people on the ground? Should "can" be analyzed as a verb or as a noun? Which of the many possible meanings of "plane" is relevant? Depending on context, "plane" could refer to, among other things, an airplane, a geometric object, or a woodworking tool."

A number of research centers at universities and major corporations continue to work on NLU advancements. In the meantime, some of the initial developments have become products. A number of firms are now involved in testing and offering NLU software. They include: Ascent Computing, BaBel Technology, Banter, Baobab Natural Dialogs, Brainhat, Cycorp, Delphes Technologies, IBM, Iphrase, Lernout & Hauspie, Leverance, Nomino, and Simplis.

### **1.3 Technology Issues**

Which technology issues impact telephony applications the most? Asked of all the key players in this market, five core issues emerged. These issues largely concern the ability of speech technologies (speech recognition, text-to-speech, and speaker verification) to work in the real world. Some of these really relate to any speech technology application including telephony ones. However, telephony environments, especially wireless services, do present special obstacles to effective speech technology implementations. Wireless environments represent some of the toughest due to noise, movement and stress on the user. The five issues described below are: accuracy, vocabulary size, natural language, standards, and network versus handset-based.

#### **1.3.1 Accuracy**

Accuracy has been a major issue in speech technology since it has been introduced. Using speech technology in consumer applications and the telephone network demands very high levels of effectiveness. This is because the current quality of landline phone communications is high. And therefore, users assume and demand high quality levels for new interfaces, such as speech, as well. According to vendor reports, current implementations of speech technology are now achieving 95 percent and above accuracy, which is what is required for the user.

One of the core issues for accuracy is the error rate. How often is someone asked to repeat what he or she says? In normal conversation, there is a certain level of redundancy. One might miss a word or not hear something correctly. We are all asked to repeat words and state things again even in normal conversations. But the level of error or restatement is fairly low; this same level is needed in automatic speech recognition environments.

The key factors that have impacted accuracy levels include vocabulary size, which is discussed below, noise reduction and filtering, and robust speech recognizers. In systems that are deployed in highly variable environments, with noise and even minimal reverberation, accuracy is likely to be lessened. Part of the solution to this problem has been that specialized components are now added to speech recognizers to deal with noise levels. A robust speech recognizer will also have the ability to self-improve. Self-improving systems can "tune" themselves as they experience and correct their efforts during trial or live applications. This self-tuning aspect allows the system to differentiate between words and just noise. As users correct the system responses, the system can learn from this and improve its recognition of accents, special pronunciations or new words.

### **1.3.1 Vocabulary Size**

The potential number of words that any speech technology system can recognize is called vocabulary size. Systems vary dramatically in the size of the vocabulary



supported depending on the application. For use as a mass market communication aid, a nearly unlimited vocabulary would be desirable. However large vocabulary systems are more prone to error because of greater potential for similar words to occur in the database. The vocabulary level that is optimal, that would combine the least margin of error with the highest vocabulary size, still needs to be pinpointed.

For simple voice dialing or voice commands applications, vocabularies of only 200-300 words may be adequate. For navigating voice portals, larger sets are necessary. Typical brokerage services support more than 80,000 words, since they need to include stock names, symbols, and nicknames. And for large national directories, nearly one million words that represent the many different surnames may be necessary in directory assistance databases. Larger and larger vocabularies can be supported as the technology improves.

### **1.3.1 Natural Language Understanding**

The ultimate goal for more sophisticated applications of speech is natural language. By accepting normal speech in sentences, the most natural of interfaces can be achieved. This allows speakers to speak without pausing or waiting for prompts. More importantly it allows the speaker to provide multiple pieces of information of their request all at once. Thus, when they ask for airline flight information, they do not have to stop and tell the system separately, where they want to leave from, what day, what time, etc. Instead they can simply say, "I want to leave Newark this Friday at 1 pm for Las Vegas and come back next Tuesday at 4 pm." In specific applications, such as requesting flight information, NLU is proving possible since the size of the vocabulary is small. As applications expand into a variety of uses, NLU will be challenged to support new levels of complexity.

### **1.3.2 VoiceXML Standards**

Voice eXtensible Markup Language (also known as VoiceXML or VXML) is a language specification for dialogs using speech recognition as input and text-to-speech or recorded audio for reply. Its aim is an open programming language used to make information content and services accessible by voice commands. The VoiceXML Forum has consolidated technologies that were independently developed by companies such as Lucent and Motorola and created a worldwide standard.

VoiceXML is the programming code driving rapid adoption of speech applications. Figure 1.5 shows how it works. In the top half of the figure is shown the actual dialogue. The situation is where a customer is calling into a company contact center. They are asked which department they want to reach. Instead of saying the one-word department name, they say, "I want billing." The system correctly pulls out the word billing and verifies it with the customer. This automatic reply is part of built-in mechanisms to speech engines that work in any dialogue. The confirmation reply from the customer is then given as yes. And the call is then routed. The lower half of the figure shows the programming code that would be used for this short dialogue. It looks similar to HTML code and works in a similar way, just that the text is not translated to text on a screen, but as speech output to the phone.

The VoiceXML Forum is the key industry consortium pushing for standardization of speech recognition tools. The Forum members are developing tools for the creation and use of this new language. It has been accepted as the model dialog markup language to be used within the W3C Speech Interface Framework by the W3C Voice Browser Working Group ([www.voiceXML.org](http://www.voiceXML.org)).

Clearly they seem to be succeeding. Over 575 companies, including Cisco, Ericsson, Hewlett-Packard, Nortel and Novell, now back the VXML specification. Using VXML, application developers can tailor specific applications for use by any VXML-compliant product. In fact, VoiceXML is poised to become the HTML of the audible Web just as HTML is for the visible Web. VoiceXML pages can now be uploaded to speech servers and hosts just as HTML pages are uploaded today. VoiceXML is going to make speech interfaces much easier to create so that it will become a widely deployed option.

The forum released the first version of VXML in March 2000. Already by October 23, 2001 the VoiceXML Forum announced it had signed a Memorandum of Understanding (MOU) with the World Wide Web Consortium (W3C). This marks the VoiceXML Forum's official release of the trademark to the public domain. The signing of the MOU also coincides with the W3C release of the first Public Working Draft of VoiceXML version 2.0.

VoiceXML is designed to work in an open, platform-independent environment. This enables everyone to participate in development of speech technology applications including network equipment firms, speech technology vendors and content and service providers. Standardization is important since it will enable mass adoption of the voice Web, along with other voice-enabled applications.

### **1.3.3 Network-Based Versus Handset-Based**

Service providers have two main choices in deploying speech technologies. They can either place speech technologies into the telephony network or add speech components to devices. For network-based versions, adjunct speech servers are added into the telephony architecture. Speech servers contain the core speech recognition engines and the applications, such as voice portals. By adding various speech technologies into the network architecture, services such as voice dialing, speech-enabled voice mail and unified communications can be offered nationwide.

Embedded speech recognition refers to the placement of speech technology within the handset or mobile device. Embedded speech enhances handset capability, making it more productive, especially in mobile settings. As devices become more multitasking units, voice-activation will ease the human interface to these tasks. PDAs and other devices that contain speech chips currently enable many speech features, the main being voice dialing and voice activation.

There are practical reasons for choosing one over the other. Handset-based solutions are much cheaper to implement and can be used with any network service.

But handset-based versions do not allow for the scalability or flexibility of network solutions. Voice dialing on handsets is cheaper but usually has a much lower limit of names; most sets come with only a few dozen up to a maximum of 200 names. Network-based Sprint PCS Voice Command service can hold hundreds of names and can be used from any existing Sprint PCS phone. There is an additional drawback to handset-based voice dialing – the directory of names and phone numbers stored on one handset cannot be transferred to another.

Examples of network-based and handset-based applications will be profiled in more detail in the next chapter. Most of the services and applications discussed in the next chapter are network-based, since the advantages of scale and deployment are necessary for many of the systemwide services, such as operator services.

## **CHAPTER II**

### **Telephony Applications and Services**

#### **2.0 Overview**

The speech technology firms, themselves, acknowledge a significant shift in focus. At the most recent SpeechTEK Annual Conference this shift was quite apparent. In many of the talks and on-site PELORUS Group interviews, it was uncovered that telephony applications are being viewed as the ones with the greatest potential. This is in contrast to the previous attention that industry players have paid to other applications.

In the past, major speech firms like Lernout & Hauspie focused on desktop dictation and voice control of desktop PCs. They sold speech software that went into desktop

software. They also sold many forms of speech technology into embedded chips that were part of machinery, toys and handheld devices. They still sell to these applications, but dictation and other forms of device control and command applications turned out to be a more limited market than the companies originally thought and much less lucrative. Today there is a shift towards network-based implementations of speech recognition that serve wireline and wireless carriers, call centers and enterprises.

The focus of this report will be on telephony applications of speech technology. The PELORUS Group interprets this broadly to include any telephone communications that can be automated or enhanced with speech technologies. As one can see from the table below, speech technology in telephony environments is finding widespread utility. The ILECs, long distance and wireless carriers have just begun to realize significant benefits of applying speech technologies to their operations and services. All the carriers we spoke to have included speech recognition ideas into their market development plans. In a few cases, they have developed and launched new services, such as wireless voice dialing, and are planning for new ones such as automated concierge services in directory assistance and voice portals in wireline and wireless environments.

**Table 2.1**  
**Speech Technology Applications in Telephony Environments**

<b><u>Application Area</u></b>	<b><u>Services/Features</u></b>
Telephone services	Voice-activated dialing Voice commands
Information services	Voice portals Wireless Web Telematics
Messaging	Voice-activated voice mail, email and unified messaging Virtual personal assistant
Operator services	Enhanced DA

	Concierge services
Security services	Speaker verification, biometrics
Enterprise telephony	Call centers, CRM Automated attendant and DA Virtual personal assistant
Mobile commerce	Voice-enabled event notification Mobile transactions

One of the services receiving the much attention today is the voice-activated information service, referred to as voice portal. In their current versions, voice portals provide voice-activated access to basic audiotext services, such as weather, travel and stock information. What once was referred to as audiotext is now being automated with voice commands. This has opened up information services to a wider market, especially the mobile population. For example, AT&T Wireless recently had a national launch for its voice portal called #121, which was created by Tellme. The potential for voice portals are now viewed as having great potential in the enterprise space too, especially for business to employee applications.

Another major service already being sold by phone companies is voice-activated dialing (VAD), which is available today from landline or wireless accounts and is built into certain mobile phone handsets. Wireless carriers have recently embraced the network-based version of the voice dialing service for their wireless subscribers. Sprint PCS has a wireless voice dialing service called Voice Command.

Unified messaging (UM) is a growing application for mobile professionals and busy householders. UM will benefit from speech technology, just as dialing has, because the interface to a subscriber's voice mail, email and faxes will be made easier. Already service providers and vendors have added voice interfaces to voice mail and email services and are expanding them to the more complex integrated UM services.

The application of speech technology in the automotive market is rapidly expanding. By adding voice activation to information services designed for travelers,

automotive firms hope to find wide application for telematics, which is defined as wireless communications for vehicles. Already GM has taken the lead by adding voice activation to its OnStar system in many of its 2001 car models.

The customer service operations of telephony firms, along with many other types of enterprises, have successfully incorporated speech technologies into their call centers. Call centers is by far the largest area of development in the enterprise space. Call centers have proved in the use of speech technologies by demonstrating improved customer service and lower costs.

Enterprises are also embracing speech technologies for basic purposes such as call switching functions and security. Companies large and small are building voice-enabled auto attendants and even virtual personal assistants for their corporate directories and messaging systems. Some corporations are taking speech technology a step further by working toward the vision of mobile commerce or electronic commerce applications that have easy to use voice-enabled interfaces. The sections below profile each of these applications and services in more detail.

## **2.1 Telephone Services and Applications**

Embedded and network-based speech technologies are already being applied in a number of areas of the public telephone network. In fact voice gateways are now interfacing to landline systems, wireless carriers, Voice over IP systems and the Internet. Figure 2.1 shows the important pivotal role that speech has in telephony services today.

The basic telephone service using speech technology is voice dialing. This allows the user to make calls by speaking a name or phone number. In addition, some services allow for call management commands, such as invoking Call Forwarding service features using spoken commands. Voice dialing and voice command services are available in wireline and wireless telephony networks. The PELORUS Group profiles applications and services in the following areas:

- Voice Dialing/Voice Command;
- Information Services; and
- DA/Operator Services.

### **2.1.1 Voice Dialing/Voice Command**

Voice activated dialing (VAD) is a good example of how advanced speech recognition capabilities have improved the application of previous technologies. Voice dialing has existed, especially on the wireline side, for some time. But an improved voice dialing service, made possible by the today's advanced technologies, offers a number of advantages over the previous versions. The current service is:

- Voice-independent -- no need to train the system to a subscriber's individual voice;
- Phone-independent -- works on any phone;
- Web-accessible -- users can create their phone directories online; they can type in names and numbers directly into a Web page form; or they can transfer them from other applications.

Thanks to speech technology, the new voice dialing service is a significant improvement over the previous versions. The system is much more accurate and flexible. Users can voice the digits of the phone number they are calling. Or if they program them into the system ahead of time, they can just speak a person's name. Names and phone numbers from a personal directory can be added to the network at a Web site. Sometimes names and numbers can be imported from personal directories stored in software like Lotus Notes or Microsoft Outlook.

With the new voice dialing services, such as Sprint PCS's Voice Command, the service is able to recognize multiple keywords and a large vocabulary. Therefore, a user can say either "call Mary at work" or "call Mary at her work number." And one can set up multiple numbers for one person. Users can say "call Helen at work" or "call Helen at home." Sprint also plans on adding a custom group-dialing list feature for



business customers. This will allow them to create directory or project-related dialing lists for their work needs, in addition to their own personal address books.

The network-based service has advantages over handset-based voice dialing. The VAD version available on handsets usually has a much lower limit of names; most sets come with only a few dozen up to a maximum of 200 names. Sprint PCS Voice Command can be used on any existing Sprint PCS phone and supports up to 2,500 names. Set-based VAD is only available on certain handsets, such as Samsung Model 2000 and some Motorola sets. And data stored on one handset cannot be transferred to another. With network-based VAD, the directory is stored in the network.

Using their voices, Sprint PCS customers can dial, look up directory names, modify their address book entries and perform various speech-enabled functions while using any Sprint PCS phone. The service costs from \$5 to \$10 per month, depending on a customer's existing rate plan. In some cases, with larger plans, the service is free. According to a company spokesperson, the residential cost is usually around \$10 per month and the business charge is usually around \$5 per month.

Nuance is the company that set up Sprint PCS' Voice Command service. Nuance is also responsible for the technology behind Qwest Wireless, who partnered with BeVocal to power new VAD and voice portal services, called Voice Calling and Voice Browsing respectively. BeVocal also helped Cingular launch its voice dialing service during November 2001. Cingular customers can buy either just voice dialing for \$4.99 per month, just voice information services for \$4.99 per month or a combined service for \$7.99 per month.

Philips Speech Processing is another speech vendor active in the VAD arena. It has sold its SpeechWave technology for voice dialing services to wireless carriers AT&T Wireless and Verizon Wireless. Through its partner Preferred Voice, it continues to sell VAD technology to a number of carriers. At the end of 2001, it extended its portfolio to several new wireless customers, including Gold State Cellular, Midwest Cellular and Cellular One of Illinois.

Voice dialing has been and continues to be an option on some of the telephone handsets that are sold by the wireless and wireline carriers. But it is by no means a common feature. Some of the handset vendors use QUALCOMM's technology. QUALCOMM offers PureVoice VR2.1, which is a handset chip-based processor that enables users to access phone functions and dialing by voice. It supports a 40-name directory.

Sensory, Inc. (Sunnyvale, California) has a processor that adds voice dialing to a number of devices including mobile phones, hands-free automobile kits, and PDAs. And Telespree, using Nuance's leading speech recognition technology, has developed a unique "keypadless" one button (on-off) handset. The completely voice-driven telephone is a prepaid mobile phone with a replaceable, combination battery pack and replenishment chip.

There are a number of other speech firms or application providers providing either network-based or device-based voice dialing including Preferred Voice, Pulsar, Comverse, Aspect, InterVoice-Brite, APEX Voice Communications and Motorola.

## **2.1.2 Information Services**

Speech-enabled information services are offered in a number of different implementations. One is as a subscription or advertising-supported voice portal for wireline customers. It can also come as a speech-enabled wireless Web service for wireless subscribers or as a telematics solution for vehicle owners.

### **2.1.2.1 Voice Portals**

Information services accessed through voice commands are called voice portals. Other names are being used as well, such as voice browsing, wireless Web and voice Web. Voice portals give subscribers access to a variety of information services through one telephone number. Speech recognition capabilities allow the user to input their

commands and navigate between content sources by voice. Text-to-speech capabilities are used to read out the data from the content sources to the user so they can hear and not have to read the information.

The current versions of voice portals are pretty simple. A user calls the voice portal phone number and listens to the available category selections. Then the user speaks the category they want and listens to the choices available within that category. They make a selection by voice and finally listen to the information, such as the current weather for their area.

Most companies that offer voice portals directly to the consumer provide a toll-free number that people call to enter the service. Sometimes the service provides a set of local phone numbers, rather than one 800 number, to use instead. The numbers, of course, can be called from either a regular wireline phone or a wireless phone. The information selections on voice portal services include a number of basic consumer information categories. The most often found categories are:

- News;
- Sports;
- Weather updates;
- Traffic alerts;
- Driving directions;
- Stock prices;
- Lottery results;
- Horoscopes;
- Flights; and
- Restaurants.

In the initial phase of the voice portal market, independent voice portal companies, such as BeVocal and Audiopoint, were offering these services nationwide directly to consumers. Since then, however, some of these offerings have been dropped or modified due to the lack of clear business models (which had been advertising revenues). Wireless carriers, though, have adopted the idea for their wireless Web

services by adding voice-activation as an interface to a subset of information content that they have already assembled for their wireless Web service customers.

All of the major wireless carriers now have an enhanced data service, called wireless Web that they offer to their subscribers. By using Web-enabled phones, users can select from a set of content listings on their WAP screens and pull up information and pictures. Wireless Application Protocol (WAP) enables tiny mobile phones to present HTML-like pages for viewing on mobile devices. All the carriers have moved ahead with WAP services first, but similar portals accessed by voice are also envisioned. By adding voice activation, the commands and entry will be done by voice. For example, Qwest Wireless has created a voice-activated version of their wireless Web service content offerings, called Voice Browsing. This service offers a small subset of Web content and makes it accessible by voice command. True browsing is not done, but rather pre-selected content in a number of consumer categories is stored and available for listening.

In the future voice portals will enable true "voice browsing", which means that using voice commands users can navigate Web sites just as they do now with search engines. Full fledged "voice browsing" is not yet available since users are not able to use word searches to browse the entire Internet. Rather, current versions allow voice-activated access to a limited set of content providers similar that is a subset of wireless Web services. By speaking one of the choices on the portals, users navigate to that specific site and use its information. By responding to the questions with spoken responses, they then navigate to the information they need. For example, they can speak the name of a company or ticker symbol to obtain a current stock price.

Since it is auditory and not visually based, full Web voice browsing must create a whole new set of voice-related functions in order to work as regular browsers. The voice browsers that are under development will try to replicate the functions of a Netscape or Internet Explorer. The voice browser, of course, will not be in the handset but on the server-side. It will conduct the functions of the user interface to the portal. For example, it will understand commands such as back and forward. It will take

search statements and execute an Internet-wide search. It will hold user preferences, such as favorite sites.

Most of the industry players expect that voice portal services will be moving to a subscription-based model rather than depending on advertisements. For example, AOL by Phone charges \$4.95 per month for use of its voice portal. InfoSpace also has customized mobile portals available with voice-activation. The major move in this market has been the shift of voice portal firms into hosting and toward enterprise solutions. Voice portal firms now create portals directed to corporate employees to help them navigate among healthcare choices and providers or within retirement benefit systems. Carriers are becoming resellers of these solutions to their large business customers.

Partnerships are key for these voice portal firms to survive. The core set of speech technologies used by these companies largely comes from two aggressive speech technology firms: Nuance Communications and SpeechWorks. These and other speech technology firms will be profiled in the Market Opportunities section below.

Table 2.2 shows some of the major voice portals and their technology partnerships. Nuance and SpeechWorks provides the speech recognition technology behind all of the active voice portals. BeVocal is out leading the pack with carrier alliances, already putting together alliances with two wireless carriers. BeVocal hosts Qwest's Voice Browsing service and is working on a trial with Sprint PCS for telematics. AT&T has invested in Tellme. Audiopoint has partnered with the speech ASP Price Interactive, which was purchased by IBasis.

**TABLE 2.2**  
**VOICE PORTALS AND THEIR TECHNOLOGY PARTNERS**

<b>Company/Product</b>	<b>Technology</b>	<b>Partnerships</b>
AOL's AOLbyPhone	SpeechWorks	Investment in SpeechWorks; Bought Quack.com
Audiopoint	SpeechWorks	Formed in partnership with PriceInteractive; iBasis bought

		PriceInteractive end-2000
BeVocal	Nuance	Begun to host services for carriers. Hosts Qwest's Voice Browsing
HeyAnita	SpeechWorks	Dialogic-hardware; Telera-ASP
Lycos Anywhere	Nuance	Run by ASP, Mobilee; Partnering with BeVocal
Tellme Networks	Nuance	Tellme studio designs VoiceXML applications. AT&T has invested in Tellme and launched its #121 service

The voice portals being developed by wireless carriers are being done so in partnership with either voice portal companies or voice ASPs. These voice portal companies or application service providers are building the content and speech recognition network infrastructure to support wireless carrier offerings.

#### **2.1.2.1.1 Voice Portals and Voice Dialing**

Many of the voice portals have also incorporated voice dialing in their services. Two the major services, AOL and Yahoo, have remote access versions of their Web site content that is accessible from any phone. These portals include the ability to dial out. AOL offers AOLbyPhone service and Yahoo has Yahoo! By Phone. In connection with its portal, Yahoo! By Phone, it offers Yahoo! Phone Card's Voice Dialing service which allows its users to make phone calls from any phone by simply saying the name of a contact in their Yahoo! Address Book or by saying a phone number.

For example, one can say "Ben Dryden at work" or "555 084 6120." This service uses voice recognition technology built into 1-800-44-YAHOO, so the user does not have to pre-record the names in advance, and they can use voice dialing from any type of telephone. All they need to do is store their contact information in their Yahoo! Address Book. The phone numbers in the Yahoo! Address Book available for voice dialing are Home, Work, Pager, Fax, Mobile, and Other. Users can store up to 5,000

names with six phone numbers each which means up to 30,000 phone numbers available from any phone.

Voice portals can also be connected to unified messaging solutions and corporate intranets. Figure 2.2 shows the expanded concept of a voice portal. It shows both the personalized home portal page on the left and the content service choices on the right. The personalized home page gives calls access to personal phone directories, calendars, etc. While accessing the same portal, users can also move to the content categories to pull out news, weather, information and other choices.

### **2.1.2.2 Telematics**

Telematics services have recently become much more available. Until recently, though, they were fairly limited in scope and were predominantly available in luxury car models. They basically consist of a single unit, built into the car at the factory or added by the dealer. The unit consists of a one-button console that links the driver to a telematics response center that provides roadside assistance, car diagnostics and maintenance alerts. Although most current units are installed at manufacture, after-market products are becoming more available. AAA has launched its telematics service center and will rely on after-market products so it can serve any car.

In its current version, telematics is an operator-assisted service. When the driver presses the call button, it dials a wireless call and connects the driver to a customer service rep located at a call center run by a telematics service provider. These reps then answer questions and provide the information needed.

Telematics services are usually a subscription-based service. The car owner has to pay a monthly service fee for the service. For example, GM's OnStar is only offered as an in-vehicle unit. If someone buys a GM car with a telematics unit they can subscribe for about \$200 per year. Although most telematics offerings work on a subscription basis, Chrysler initiated a new service and pricing plan in October 2001, which is pay-per-use structure, instead of subscription. It will be made available in spring 2002.

There is a new urgency for hands-free phone calls and information services in cars because of lawsuits and new regulation. New York State actually passed a law against driving a car while using a mobile handset. Some 40 states have considered or passed laws related to vehicle use of wireless devices. In the recent past many of the basic telematics services did not include the ability to make personal voice calls. Now many cars come with hands-free phones built in. Various BMW, Jaguar, Lexus and Mercedes-Benz models come with special versions of Motorola or Denso phones. These are linked to the car's radio and to a built-in microphone near the driver. With the phone in its cradle, the driver can make and take calls without moving his/her hands from the wheel.

Fortunately, with the improvement of speech technologies, telematics firms are now able to add voice-enabled capabilities. In some cases, these systems even eliminate the live customer service rep and allow the driver to interact directly with the information, just as it works with voice portals. In the future, telematics services will deliver hands-free wireless calls, unified messaging and voice-activated Web access.

Here is how a basic telematics system works:

- To initiate a call, the driver presses a call button to establish a cellular connection with a telematics service bureau server;
- The driver can ask questions and get help from the telematics rep;
- In some cases a network server receives the vehicle's location data from an on-board global positioning system receiver;
- The driver asks for directions to a specific location, such as the closest location of a restaurant or bank;
- After the request is processed, a live rep or a computer-generated voice broadcasts the results over the vehicle's audio speakers.

Telematics is actually envisioned as a fairly complicated and comprehensive set of automated diagnostic, notification and communication services for vehicles. This includes driver information services, such as travel directions, and engine management systems, such as automatic diagnostics. Telematics is referred to by Sprint PCS as



“concierge in the automobile.” It offers critical travel information needed by consumers and business people on the move. Figure 2.3 shows how telematics is envisioned in the future. It shows both voice input and keypad entry of requests to the dashboard console. This leads the user to a number of uses, including alarm, GPS location, hands-free radio use and telephone calls, and various data services through the onboard data device.

Before the year 2000, telematics had largely benefited the trucking industry. Segments of the trucking industry have been tapping into wireless communications via the data networks of QUALCOMM’s Omnitrac, for example. Now though, the Big 3 automakers have installed wireless connections in cars that offer basic security and roadside assistance. The big 3 automotive firms have made various partnerships or set-up their own divisions to offer telematics. General Motors (GM) has its OnStar subsidiary. Ford’s automotive parts spin-off, called Visteon, is also offering telematics. Ford has also formed a partnership with a leader in commercial trucking communications, QUALCOMM. DaimlerChrysler is using other independent telematics firms, like ATX, in its plans. In fact there are a number of telematics service providers, such as ATX Technologies, Webraska Mobile Technologies and Acunia, that have packaged together software, content, wireless, and customer services into services that are branded by the automotive firms. Telematics device makers include: Bosch, VDO CC, Motorola, Delphi Automotive Systems, and Visteon Corporation.

### **2.1.3 Operator Services**

Another application area for speech-enabled telephone service is operator services. Automation in operator service centers already depends on speech technology. The major focus in this area is directory assistance (DA) service. Speech technology is being implemented into basic DA, enhanced DA, as well as incorporated into enterprise DA solutions.

### 2.1.3.1 Directory Assistance

Some of the largest early implementations of speech technology were in the basic operator services arena. In early 1992, AT&T began to implement a program to replace about one-third of its long-distance operators with ASR technology. This program still represents one of the largest ASR implementations anywhere. AT&T was able to keep it simple and used speech recognition with a vocabulary of only five words to automate small portions of operator-assisted calls. This service generates over a billion speech recognition attempts per year and saves AT&T an estimated \$200 million per year.

This is how it works. The directory service applications allow users to request residential or business listings by simply responding to questions related to their request. One example of an automated DA call flow is seen in Figure 2.4. This example from varetis, a Munich-based provider of directory solutions, shows how the computer and caller interact. For example, a caller dials a number to get a residential listing and hears the prompt, "City and state, please." Once the caller responds, they are prompted with "What is the last name?" This works through until the number is provided.

Various DA providers offer platforms that include speech enhancements. The DA provider, ISx, for example offers irectory assistance through its Automated Operator Assistance Platform. The platform includes:

- Recording the inquiry;
- Analysis by the speech recognition engine;
- Generation of the database inquiry; and
- Audio delivery of the listing information.

For requests with DA operator involvement, the operator listens to audio playback of the caller's request with the workstation already displaying possible listings. When only the locality is recognized, the search screen is delivered with locality pre-populated, along with the recorded inquiry. The base application, Voice Store and

Forward, prompts the caller automatically, and the request is recorded. The silences are trimmed for playback to the operator. Prompts and systems options can be customized for each specific market. The ISx Intelligent Audio Subsystem automates the delivery of requested listings. After the operator selects the correct listing, the information is released for audio delivery to the caller. SMS delivery of the listing data to a wireless handset is also possible.

Using automated DA, after the listing is found, the system responds with the requested number and gives the caller a choice of one or more options. They can hear the number or have number dialed for them. The later function is a service called Directory Assistance Call Completion or DACC. In voice portal services, such as InfoSpace services, users can choose to complete the call, save the listing to their personal address book or receive the listing via SMS on their mobile phone text screen.

What are some of the firms operating in the DA space who have incorporated speech technology? Phonetic Systems is covered in our prime speech technology firms, since it has its own speech technology. ISx, varetis and VoltDelta are profiled in the best-of-breed section in Chapter III.

### **2.1.3.2 Enhanced DA**

Speech technology can also be effectively applied to enhanced DA services (EDA). Speech technology is particularly relevant to EDA services, which include any services beyond the basic look-up of local telephone numbers from a name and an address request. These include a range of services.

The automated EDA systems work in slightly different ways. As one example, the varetis system is seen in Figure 2.5. It shows how the automated EDA system fits into the telephony network. Database interfaces are necessary to control various search engines for the National DA, concierge or location-based services as described below. Fallback operators are always available to barge in when the caller has any problems.

SMS messages may be used to deliver the directory listings or any information delivered from EDA applications.

### **2.1.3.2.1 National Directory Assistance**

One of the most common EDA services is national listing, often referred to as NDA (National Directory Assistance). It is offered by nearly all major Directory Assistance providers, including local telephone companies (ILECs, including the former RBOCs), long distance carriers, and major wireless providers. The real challenge in this area is to incorporate speech recognition into the major National Directory Assistance (NDA) databases. This is a challenge given the enormous number of listings. The barrier in the past has been the size of these databases because no speech vendor has been able to handle such large vocabularies. A handful of vendors are trying to tackle this challenge, with major efforts by Nuance and Phonetic Systems.

Nuance claims that it is close to offering a partial solution. Since 80 percent of Directory Assistance calls are for business and government listings, and these listings are much smaller in number, at least this set of numbers can be automated. This is the approach taken by most so far. Phonetic Systems, for example, chose to focus on business numbers because about three-quarters of all 411 calls are to get business numbers. Also, it is easier to accurately decipher names of businesses and provide accurate numbers than it is for millions of individual people, especially people with unusual names. Phonetics is trailing the service. Working with LSSI, a New Jersey-based information company that maintains a database of 175 million US phone numbers, Phonetic has assembled listings of 700,000 business numbers in the four Eastern Massachusetts area codes for its "Free D.A." number.

To solve the larger problem, Nuance has further demonstrated a system that aims at voice-command nationwide DA supporting up to 100 million listings. They say this should automate about 75 percent of total Directory Assistance calls. Phonetic Systems has developed a phonetically-based system that offers a DA service called Directory-

Assistant. Directory-Assistant has the capacity, according to Phonetic Systems, to work with databases of more than 500 million listings. This is done through sophisticated initial screening that brings the request down to a manageable size.

#### **2.1.3.2.2 DACC**

Besides NDA, voice-activated DA offers Directory Assistance Call Completion (DACC), which immediately connects callers to the number they requested from Directory Assistance. Two other EDA services that use the NDA database are Reverse Look-Up, which enables a caller to get a customer name and/or address from a telephone number, and Yellow Page Searches, where a caller can request a type of business, such as a plumber, without having a specific business name. The DA operator then provides the name and number of a business, or several businesses, for that category.

#### **2.1.3.2.3 Concierge**

The other category of EDA services utilizes information from data sources other than DA databases. These services are sometime referred to as "concierge" services. There are a variety of services that DA providers can offer, such as traffic, weather, directions, movie listings, restaurant reviews and reservations, stock quotes, and sport scores.

Companies offering these services need to be able to access accurate, timely information to provide up to date information. Many of these types of services rely on information that is dynamic -- changing on an hourly, daily, weekly, or monthly basis. While traditional DA database information changes on a daily basis, most the specific information is constant for years. This is not the case with services such as traffic conditions and weather. These services require access to real time information to be useful.

During the last year or two, speech recognition technology has improved so much that new information services based on voice-activated platforms have been created. EDA sites can be speech-enabled so that callers from any phone can call in, listen to basic commands, reply with their voice and then listen to the information read out to them. Voice portals have also expanded the reach of EDA services.

Ultimately, operator services are attempting to develop the concierge services into natural language implementations. Carriers have been experimenting with a service that can start simply as "How May I Help You?" where a customer can say anything he or she wants and the computer carries on an almost human-like conversation to address and pinpoint the users needs.

### **2.1.3.3 Location-Based Services**

For now, much of the commercial application of location-finding technology relates to DA services. EDA providers are linking with firms such as InfoSpace and go2Systems, to link directory data to location-data such as store location and street directions. Once E911 deployments take place, which will be soon, further commercial location-based services will emerge. Figure 2.6 shows how location-based services will be implemented. Signals from the mobile device or network will translate precise location via the location platform. This platform will in turn facilitate interaction with the various application sets, such as E911, location-sensitive billing, etc. For the commercial services, content providers will provide the link to the various information sets, such as shop locations and traffic information.

When fully realized location-based services will use automated location identification techniques to pinpoint the location of a mobile user. This data would be linked to directory data so that a user can call up and find out local retail sites. For example, they would search for the nearest gas station and get street-by-street directions to find it. Current location-based services rely on manual input. In other words, the user inputs their location and then the service delivers relevant data.

InfoSpace and go2Systems are both active in this area. They have successfully tied Web directories to mobile user location turning the data into useful location-based services.

Given recent governmental mandates, such as E911 in the US, vendors are implementing the location-finding technologies and in some cases linking it with speech technology. One example is Nortel Networks. Nortel and wireless location technology provider, Cell-Loc Inc. field tested a voice-activated location-based service for mobile phone users. Cell-Loc worked with Nortel last spring in Austin, Texas, to jointly test the two companies' latest wireless service geared for existing and future high-speed 3G networks. In addition to operating on traditional cell phones, Cell-Loc said the new location-based service would be accessible via personal digital assistants (PDAs) and Web-enabled mobile phones.

In the service, wireless users are able to access location-sensitive Directory Assistance and navigational information without providing an address, cross street, zip code or city/state combination. A caller's location is determined automatically using Cell-Loc's Cellocate TDOA (time difference of arrival) location network. Nortel's voice navigation system translates the caller's voice request into data, enabling location-specific services -- such as searches for nearby restaurants, shops or other businesses - - through its National Directory Assistance Service. Search results are converted into a voice response that is played back on the user's mobile handset. Trial participants are able to create a custom profile on a designated Web site. (To find out more about the location-based services market, see The PELORUS Group's new report, Location-Based Services Market, March 2002)

## **2.2 Enterprise Telephony Applications**

Many consider enterprise applications as the best path to near-term revenue potential for speech technology firms. The current application of speech technology by

enterprises range from systems which cut costs in calls centers to improved contact interfaces that enhance customer satisfaction and company image.

Figure 2.7 shows that speech technology can be at the heart of new automation platforms that impact critical groups and entry points into the enterprise. Enterprise automation systems connect legacy corporate systems with critical users groups -- customers, mobile workforce and on-site specialized knowledge workers. It does so in order to improve accessibility to various critical functions, shown as the triangles around the core enterprise automation platform. These include access and improved interfaces to: corporate Web sites, ATMs and kiosks, email and voice mail systems, and corporate wireless networks. Contact centers often act as the facilitation center to these systems and provide entry into the core data sets and capabilities of the corporate resources, such as legacy systems, databases and telecom networks.

Speech technologies have become viable alternatives to the longstanding use of interactive voice response (IVR) systems. Speech recognition, text-to-speech and voice authentication technologies are being used in a variety of ways in major corporations and now they are moving down into mid-size and small enterprises. This is especially true for call centers and information portals, where voice-activated interactions are used to create customer self-service, reduce costs, and generate new revenues. Call centers are finding that it is possible to introduce a much greater range of functionality than was available through traditional touchtone IVR. Speech-enabled applications allow callers to quickly progress through options using spoken commands or even natural language rather than tediously progressing through long touchtone menu selections.

Speech recognition is gaining acceptance in call centers as a self-service tool. Customer self-service means that the customer can use the automated system to find the status of their account or to extract information from a database. In fact, customer self-service has progressed and is beginning to be used for even more complex transactions. As speech systems become more sophisticated, they will enable more self-service transactions. With more of the traffic offloaded, live agents will be able to



concentrate on answering more complex inquiries from customers, to focus on cross-selling and to conduct up-selling.

The number of different enterprise speech applications is almost limitless. Voice-enabled applications for enterprises comprise many different operational and functional areas. The table below shows one view of these. The table includes applications that are already offered or being designed to be enhanced with speech technology:

**Table 2.3**  
**Speech-Enabled Enterprise Applications and Functions**

<b>Application Areas</b>	<b>Functions</b>
Customer service	Customer service inquiries Order status & inquiry Appointment scheduling Direct voice to text Retail order tracking Customer care application Other CRM
Sales force automation	Lead tracking Field rep access to company data Sales data for reps Sales database applications
Human resources	Employee benefits self-service HR enrollment Benefits management New hires, transfers, exits
Financial services	Bank transactions Brokerage transactions
Personal assistant	Voice dialing directories

	<p>Conferencing by voice</p> <p>PIM</p> <p>PDA wireless interface</p> <p>Personal schedule/message center</p> <p>Directory for call routing</p> <p>Access to calendar</p> <p>Personal mailboxes</p> <p>Time and response reporting</p> <p>Other UM enhancements</p>
Travel support	<p>In-vehicle telematics</p> <p>Location-based services</p> <p>Help me with directions</p> <p>Vacation information</p> <p>Shops and local information</p> <p>Personal location directions</p> <p>Airline flights</p> <p>Traffic alerts</p> <p>Severe weather alerts</p>
Ordering	<p>Telephone, catalog ordering</p> <p>Self-service applications</p>
Field service	<p>Field service technician support</p> <p>Field service management support</p>
Corporate data access	<p>Database access</p> <p>Network data access</p> <p>Enterprise voice portals</p>
Training	<p>Sales training</p> <p>Registration, billing, tracking, certification processes</p>

Procurement	Replenishment for existing customers Auction notification Industrial inventory Business-to-consumer procurement Frequent order purchases
-------------	--

As seen from this table, it is important to narrow the focus in this overview report to the essential areas. The PELORUS group has discussed this issue of focus with key players and has distilled the critical applications to focus on in the near term, which are:

- Attendant services (operator) -- automated attendant functions related to dialing extensions and directory/database searches;
- Messaging -- voice mail, unified messaging and virtual personal assistant enhancements; and
- Contact centers (agent) -- IVR functions related to call routing (auto attendant) and customer self-service (portals).

### **2.2.1 Attendant Services**

The basic application of speech technology in the enterprise is for incoming call handling. Businesses can use speech technology applications to partially or completely automate the routing of incoming calls. This is referred to in a number of ways, either as call routing, automated attendant or Directory Assistance. An automatic attendant replaces a live operator, playing recorded prompts to handle incoming telephone calls.

Its basic function is to automate the operator or switching function for incoming, internal or outgoing calls. Automated call routing can be accomplished by retrieving information before the caller is connected to a person. For example, an automated attendant system can answer an incoming call and query the caller through a series of

question prompts. The caller can give their replies through speech-activated responses. This dialog can determine the destination department or person that is requested and the call can then be transferred. This is especially useful as a means to back up live operators during busy periods of the day or to route callers to voice mail systems at night or on the weekends.

However, while keypad touchtone has been the normal means for interacting with auto attendants, speech recognition is becoming more widespread as the means for interfacing. The speech-enabled version of the automatic attendant lets callers respond to the prompts using natural speech instead of touchtone digits. By allowing users to speak their requests rather than respond by keypad, it provides a more natural and comfortable user interface. Speech recognition provides some key benefits including faster call handling and shorter menus so that callers can simply speak the names or departments they want. Vendors point to satisfied customers who have found that they have achieved an ease of use that results in higher call completion rates and hands-free input for callers.

Either sold as an adjunct to a PBX or as a feature of voice messaging systems, auto attendants have proven valuable to companies for both after-hours attendant service and overflow coverage when company operators are busy. Auto attendants provide a way for first-time callers to find a person in a company, without requiring callers to know the direct-inward-dial number of the person being called. Auto attendants perform routine call answer and routing functions, while some of the more complex versions provide further integration with applications such as interactive voice response (IVR), voice messaging and other functions.

Auto attendants are not just used by large corporations, but have made their way down to even mid-size and small enterprises. Smaller companies deploy them because auto attendants are a way to help them appear as technologically sophisticated as their larger competitors. Plus by using the automation, it eliminates the need for office managers or administrative assistants to answer incoming phone calls.

Callers within an enterprise usually generate more than half of the call traffic handled by corporate operators. With a speech-driven auto attendant, an employee can pick up the phone and speak the name of the employee he or she wants to reach without having to look up or ask for a phone number. This has cost saving benefits, since it helps eliminate personnel costs as well as the cost to print corporate phone books. In larger enterprises, the savings on printing and distribution alone can be substantial. Combined with lower personnel costs, the return on investment is attractive.

There are a number of standard features that are usually part of auto attendant systems. The basic feature is the ability for routing via menus, which allow users to voice navigate through multilevel menus similar to IVR ones. When speech technology is deployed it is usually as an overlay over existing IVR systems, so that the resulting systems are dual-use, which lets a caller use speech or DTMF input as desired. The other key features are:

- Barge-in, which allows the caller to interrupt the auto-attendant prompts and speak a request;
- Scheduling greetings by time of day, day of week or holidays;
- Call transfer;
- Text-to-speech for either reading unfamiliar names in the database or for reading information from audiotext or Web sites;
- Fax integration for access to the company's fax number;
- Natural language understanding which enables spoken sentences as input;
- Automatic directory updates can be done in real-time while the system is still up and running;
- Access to outbound calls using personal directories, contact managers or PIMs;
- Voice-activated dialing allows for calling to external numbers, pagers or other numbers through voice commands;
- Routing based on Caller ID or ANI; and

- Voice verification used for password protection for PIM information, access to personal directories or access to the telephone features of the PBX.

### **2.2.1.1 Auto Attendant Features**

Many of the systems permit some level of voice messaging integration, as opposed to a simple transfer of a call to the voice messaging system or mailbox. This feature lets the system make use of digital information that comes with the call. For example, speech recognition might be used to extract a caller's phone number from a voice message for callback. Multiple language capabilities are also available, mostly for the main European and Asian languages. Some of the languages included in the automated systems are: French, Spanish, German, Italian, Portuguese, Finnish, Dutch, Castilian Spanish, Chinese, and Japanese.

As with voice processing systems, auto attendants come in varying port sizes. A few vendors, such as Phonetic Systems, specialize in large-vocabulary enterprise applications, with systems of 192 ports (which equals eight T-1s) and above. Other vendors, such as Sound Advantage with its SANDi product, can also accommodate large amounts of users through networking systems together. However, the majority of systems are sold in the 12 to 24-port range.

The other key measurement is related to the number of names that can be handled, which is referred to as vocabulary. For vocabulary size, a base number of 20,000 entries is considered the norm. Searches are done in flat file databases that treat each word or name as a separate item. To handle very large searches, of more than 1 million, screening of the request is done to drill down into the database to reach a smaller subset of names for searching. Three vendors, Lyrix, Parlance and Phonetic Systems, claim the capacity of a million records or more, which is easy to reach if large companies or public phone lists are included.

A range of solutions relating to PBX integration with auto attendants are also available. Some vendors offer digital telephony integration, including Locus Dialogue,

Mitel, Phonetic Systems and Sound Advantage. Digital integration lets vendors offer features, such as routing based on Caller ID, that are not possible with pure analog systems. Locus Dialogue, Parlance and Sound Advantage all publish a long list of possible digital and analog auto attendant integrations for popular PBX vendors, such as Mitel and Nortel.

Vendors offer a variety of methods to handle system administration, end user set-up, remote administration and directory updates. Remote administration is fairly standard across all vendors. Similarly, most vendors allow their end users to set up personal call lists for frequently called numbers. Some vendors make the end users go through the system administrator to make changes, while others allow users to access their own listings through a Web-based administration screen.

Auto attendant firms include: Phonetic Systems, Sound Advantage, Lyrix, Parlance, Locus Dialogue, Mitel, ONEder Technologies, and Artifsoft.

## **2.2.2 Messaging Services**

Speech technology is also being deployed to create a new interface to all types of messaging services. This includes everything from adding voice controls to basic voice mail systems to speech-enabled unified communications. Email systems are also taking advantage of text-to-speech improvements so that emails may be retrieved, even on cell phones, and then read over the phone to the user.

### **2.2.2.1 Voice Mail and Email**

Voice mail and electronic mail systems have become complex communication tools for everyday business and a critical necessity for mobile employees. In today's sophisticated voice mail systems, one can easily access messages, make selections, listen to them and change system options from the road. A callback feature enables automatic callback to the individual who left a message. Some applications allow group

mailbox lists to be created for sending the same message to multiple recipients. Mailboxes can be set up that are one-way "listen only" -- designed for organizations that need to make special announcements. One example would be power companies that must warn of outages during storms. Voice mail has been expanded to include auto attendant and follow-me capabilities. All of these features have made the interface much more complicated, though. This has driven up demand for use of voice-activated interfaces.

The same principle applies to electronic mail. Email store-and-forward systems have become fairly sophisticated. Since email is so essential to conducting business today, it must be made accessible from everywhere. Voice-enabled email uses voice recognition and speech synthesis technologies to enable users to access their email from any telephone.

In general, the various speech-enabled messaging products work the same. The subscriber dials a phone number to access a voice portal, from which they collect their email and/or voice mail messages. This is done simply by pressing a couple of keys or saying a phrase like "Get my email." Speech synthesis software converts email text to a voice message, which is played back over the phone. The user may navigate through options (such as skipping messages or hearing a list of senders) through voice commands or keypad selections. Users can even dictate their replies, which are delivered to the recipients as voice messages. Voice-enabled email is especially useful for mobile workers, because it makes it possible for them to access their messages easily from virtually anywhere without having to invest in expensive equipment such as laptop computers or personal digital assistants.

One of the key offerings in this space is Comverse's Voice-Controlled Email (VCEM) System, which is an email management solution. VCEM's voice activated interface enables users to compose and respond to email messages verbally. A personal address book stores multiple entries, which enables users to verbally reply by email or dial the sender by simply saying "Return the call." VCEM works with a wide range of handsets, for both mobile and wireline phone systems.



Comverse has deployed voice-enabled mobile messaging service in Blu S.p.A., the fourth-largest wireless operator in Italy. Blu's wireless service enables users to control their voice mail with spoken commands. Users are able to review, respond to and delete messages just as they would use a traditional keypad. In this deployment, controlling voice mail by voice is the default choice of the system, so that users automatically access their voice mail by spoken commands unless they decide to use the keypad. Comverse conducted research to ensure maximum quality and reliability of its voice-controlled messaging service. This entailed collecting sample speech patterns from various regions of Italy. Using this information, Comverse created a speaker-independent, voice-controlled solution accommodating commands that would respond to various Italian dialects and accents.

Philips Speech Processing, for one, is convinced voice-activated access to voice mail is a critical speech technology application today. They say they are seeing thousands of ports added by their customers in the move toward making voice mail mobile-friendly. Currently, Philips is working with various ASPs to help some 15 different carriers upgrade their voice mail systems.

Wildfire Communications is one of the personal assistant/unified messaging suppliers that is taking advantage of speech technology. Wildfire has built a combined voice dialing and voice mail application that is attractive. In Wildfire Personal Assistant, the basic service allows the user voice dialing and voice-activated voice mail. Wildfire sells the wireless carrier the software, which runs on a Compaq Proliant server platform, and the carrier manages the system. Wildfire loads the software onto the carrier's server and integrates it into the wireless carrier's network.

SpeechMail, a system developed by Vocalis, allows email users to access their emails solely over the telephone. It uses a combination of advanced speech recognition and text-to-speech technologies. SpeechMail is already widely used by many leading ISPs, including BT Internet, Cable & Wireless and UUNET.

#### **2.2.2.2 Unified Messaging**

Unified messaging integrates a number of capabilities into one central access point so that users have access to voice mail, email, fax and pager messages from any device. This central box means that speech technology can be effectively employed for users to control all of their messages by voice command. This can be done from whatever device they are using, whether it be wireline phones, mobile phones, or personal digital assistants (PDAs). Unified messaging solutions are available across a broad range of architectures, including wireline, wireless and enterprise telephony systems.

Unified messaging works by integrating various forms of communications into one interface, usually the phone. Two examples of fee-based UM services are LinxConnect and TelePost. Linx Communications provides LinxConnect, which offers UM through a single telephone number that belongs to the subscriber. With LinxConnect, the user:

- Gives out the single number to all their contacts;
- Sets up a follow-me route for incoming calls;
- When calls come into the single number it forwards the call based on follow-me routes prescribed by the user;
- If the subscriber cannot be reached, the systems routes the call to a voice mailbox;
- If a fax comes in, it is automatically sent to the fax mailbox.

In addition to this basic service, TelePost offers conference calling, contact manager and ring-me-now features. And TelePost offers email message retrieval.

Typical applications of UM are found in larger enterprises with the purpose of integrating UM functionality into their existing voice mail systems. Companies can also buy separate UM systems or use hosted services. Enterprise UM providers include Avaya, Nortel and Cisco. Hosting services come from Cingular Interactive, CyberTel, Webley Systems, or Openwave Systems. The most prevalent equipment deployments of UM are as adjuncts to conventional PBX voice mail deployments. Sometimes the systems have links to LANs and other messaging servers such as Microsoft Outlook and

Lotus Notes. Most voice mail vendors offer or are developing unified messaging upgrades to their corporate office solutions.

The use of UM is expanding to a host of other markets besides large enterprises. As noted in The PELORUS Group's recent report on unified messaging (see [www.pelorus-group.com](http://www.pelorus-group.com) for the report, Unified Communications for the Service Provider Market), application service providers are fueling the rise in the use of unified messaging. Individual consumers or small companies can take a subscription to UM services from providers such as Ureach (Holmdel, NJ), etrieve (Hillsboro, OR) and j2 Global Communications (Los Angeles, CA). They pay from \$10 to \$50 per month for a toll-free access number, online storage space for faxes, email and voice mail messages, Call Forwarding features, plus a personal Web page. Additional features include conference calling and international calling. ASPs believe that adoption of WAP phones will go far toward broadening the market for unified messaging into the wider SOHO and consumer markets.

Until now, one of the significant barriers to wider adoption of UM has been the need for a simpler user interface. With voice recognition interfaces, the solution is here. Speech recognition is being applied to unified messaging functions in a number of ways, from the most basic to the more complex. They include:

- Voice-controlled voice mail -- controls voice mail features, such as playing and saving messages by voice command;
- Voice attendant -- voice-activated call router (auto attendant) which directs calls by speaking the name or department desired;
- Voice feature control -- controlling network-based enhanced features such as call forwarding or conference calling by voice commands;
- Personal assistant -- an integrated set of features, such as voice mail, call screening, voice dialing and call forwarding, all controlled by one consistent voice interface; and

- Voice-controlled unified messaging -- control of all the features of unified messaging with spoken commands and text-to-speech synthesis to read email or fax headers.

In the last two years there have been several announcements by unified communications providers explaining that they have speech-enabled their interfaces. UM firms have formed strategic partnerships with speech technology firms and have added automatic speech recognition and text-to-speech technologies in combination with existing interfaces. One was the ability to convert email text into audio messages via text-to-speech so that it could be delivered over the phone. Until now, however, voice mail systems used touchtone input for control functions in conjunction with pre-recorded prompts and system responses. With speech-enabled input and voice responses, the user interface can become hands-free.

Major UM vendors, like Nortel, are beginning to integrate speech recognition into their systems. The UM system, called CallPilot, has built-in speech recognition that provides voice interfaces to voice mail and fax messages. The capability is now being expanded for voice access to email. Another example is InTouch Systems (Cambridge, Massachusetts), which has partnered with IBM to develop voice command remote communications for mobile workers. Using IBM's ViaVoice technologies and Intouch's software platform, called InFlexion, mobile employees can access voice mail, email, faxes and calendar data.

A number of speech vendors offer TTS that could be used for UM applications. This includes Lernout & Hauspie, SpeechWorks, and Lucent. Table 2.4 shows the subset of UM firms which are involved in adopting speech recognition into their UM platform or UM service offerings. This includes major network equipment vendors, like Nortel Networks, and speech firms, like Philips. It also involves specialized firms, like Vocalis and Tornado, along with innovative UM service firms, like ThinkLink.

**TABLE 2.4**  
**UNIFIED MESSAGING PROVIDERS**

<b>Company</b>	<b>Description</b>	<b>Speech Recognition</b>
----------------	--------------------	---------------------------

eVoice	UM services provider	Offer carriers and customers voice-enabled messaging services
InTouch Systems	Partnership with IBM	Voice command remote communications for mobile workers
Nortel Networks	UM Vendor	Call Pilot, a unified messaging system, has built-in speech recognition capabilities
Philips Speech Processing	Speech technology firm	Voice activated voice mail
Shoutmail	Voice portal firm	Messaging voice portal
Tornado	UM technology provider -- UM engine called Tornado Messenger	Has added speech recognition to its UM engine; partnered with SpeechWorks
Toshiba	Platform vendor -- network server, called Strategy Enterprise Server	Offers new Messaging Voice Processing Application that adds voice-processing features
Vocalis	UM and speech technology firm	SpeechMail is voice-activated email system
Webley	UM company	Creating voice-activated personal assistants
Wildfire	UM services provider, branded as Wildfire	Has deployed voice interface to its personal assistant w/Orange plc, the UK wireless firm

Combining speech recognition with personal assistants and UM has been gaining popularity, as it allows hands-free operation of the UM system, while enabling users to be more in control of their messaging and communications.

According to Virtual Plus, the next big thing is Speech to Text -- i.e. recognizing a person's voice and turning that into text. At the moment simple phrases can be recognized -- such as "get messages" or "list appointments on Monday afternoon" but within the next two years, we will be able to command the computer for a full range retrieval functions (for example "between 12 and 2 direct my faxes to this number"). Not only that, but the computer will be able to "hear" a regular human voice and turn

that speech into text. Once that message is “texted” it could be delivered via a number of means, such as being emailed to the recipient or sent as an SMS to a mobile phone, or even sent to a fax machine. The user could dictate their message for delivery via a variety of mediums.

### **2.2.3 Contact Centers**

Service bureaus and enterprise call centers are also benefiting from speech technologies. The travel and hospitality, catalog, package delivery, financial services, entertainment and ticketing industries have already adopted various speech technologies into their operations. The key advantage of adopting speech technology is the ability to increase capacity, increasing the number of calls and transactions handled. It also allows for a decrease in the number of live operators required to handle the load.

Traditional IVR has met the need for contact centers for many years, but now with speech technology available it is viewed as inadequate. Contact centers report that their users feel that traditional touchtone IVR is inconvenient and at times bothersome. Bypassing the required prompts of press X for service, press Y for sales, will speed the calls along and create higher levels of customer satisfaction.

Already a number of industries, especially travel, have adopted the new automated speech recognition interfaces. The contact centers for major airlines already use speech technology. American Airlines uses Nuance’s speech recognition and United Airlines uses SpeechWorks’ speech recognition to power their toll-free response lines.

System integrators and VARs are also selling into call centers. This is especially true for customer relationship management (CRM) applications. Gold Systems supports multiple IVR platforms and speech recognition technologies, including Nuance, Philips, and SpeechWorks. It supports platform partners: Aspect, Avaya, Cisco, HeyAnita, InterVoice-Brite, and Tellme. According to industry research, users report that the most important initiative for CRM efforts is self-service. CRM software includes sales

force automation, field service, and contact center applications. Contact information, calendars, sales opportunities reports and task lists can be accessed and updated in near real-time basis using the telephone. Business-to-business Web sites can be speech-enabled to provide service to online customers. This replaces IVR interfaces for technical support requests and access by suppliers and customers to corporate knowledge databases.

PeopleSoft is one of the large CRM vendors with a speech technology strategy. According to its recent white paper (The Case for Speech-Interactive Enterprise Applications), PeopleSoft Mobile CRM is its strategy for serving the needs of on-the-go sales and technical personnel. Using wireless devices, those in the field can connect to corporate databases, including back office and transactional systems. But building an effective user interface has been a challenge given the multitude of device types and CRM applications in use. Using speech technology some of the interfacing problems can be solved. PeopleSoft has partnered with JustTalk to deliver natural language interfacing to its CRM systems. PeopleSoft CRM functionalities, like task list and calendar management, are combined with JustTalk's Speech Solution to create speech-enabled applications.

### **2.2.3.1 Virtual Personal Assistants**

Virtual Personal Assistants (VPA) may be part of basic enterprise operator functions or as part of call centers. VPAs involve a set of applications that act as personal secretaries to the subscribers. They provide "friendly" automated assistants that create remote access to address books, network-based calendars and voice mail. Network features like conferencing, call return, and call forwarding are also sometimes incorporated into the VPA system. Some of the leading personal assistants are Wildfire, Webley, and General Magic's Portico.

Virtual Personal Assistants that use speech recognition to help manage messaging and communications will be increasingly important as people become increasingly

reliant on messaging traffic. Personal assistants perform a range of services on behalf of a user when they call in by understanding the commands they are given. Personal assistants can be considered a layer on top of unified messaging. PVAs can respond to spoken commands such as: "How many new messages do I have from my boss", "What are the subjects", "Read me the second one", "delete it", and so on.

For example, Cisco Personal Assistant works with Cisco CallManager and is designed for the busy mobile executive. Users can access voice mail, dial by name and conference from any mobile phone using voice commands. The phone services enables the user to check email, voice mail, calendar, and personal contact information.

Use of virtual assistants will enable whole new forms of customer care. Using virtual assistants in customer relationship management (CRM) will permit higher levels of customer self-service. Already firms such as General Magic, Nuance and SpeechWorks have developed suites of solutions and rapid prototyping environments for call center and CRM applications.

The following is a scenario from a Microsoft white paper on unified messaging depicting the use of VPA and speech technology:

"Tom, a national sales manager, has a short layover at O'Hare. His virtual personal assistant (VPA) calls him on his mobile phone, having determined from Tom's calendar that he should be between planes. The VPA immediately announces two critical messages -- that is, messages from someone on a list Tom maintains or messages containing any key word he has included on the list. Tom tells his VPA, using natural language, to read the messages to him, which the VPA does, even though they are in different media.

An important message that has been expected from his boss has not arrived. But he doesn't give it a moment's thought because he has a standing order in his VPA to find him and notify him as soon as any message from his boss arrives. The VPA will attempt to reach him through a progressive set addresses depending on the day and hour. If



it's a workday between 8 a.m. and 6 p.m. and Tom's not traveling, his VPA will "ring" Tom's desktop PC. If Tom doesn't respond, the VPA will try Tom's mobile phone, then his pager, then, by Tom's direction, every other number and address he has. If it's outside working hours or on the weekend, the VPA calls Tom's home phone first. Between midnight and 6 a.m., the VPA doesn't call at all unless the boss' message is tagged urgent. When Tom's boss finally sends the message, Tom gets it within minutes and responds immediately."

As an extension to self-service and portal services, call centers are also offering virtual personal assistants. Virtual personal assistants allow the contact center to provide much more than the normal request for information. Designed to take requests in natural language, they can handle fairly complex tasks. For example, someone can call an airline number and ask a virtual assistant about airline flights. They can ask the assistant to perform a task, as complex as creating an alert so that they can learn of the impending arrival of a family member's flight so they can go and pick them up. The user of this VPA might call the system and state something like, "Check on American Airlines flights from Houston to Boston, leaving Houston about 1 PM. Call me one-half hour before it actually lands in Boston."

### 2.3 Conclusion

The diversity of applications and potential services that utilize the capabilities of speech technology is staggering. Speech technology impacts even the most basic of telephony functions, such as using one's voice to "dial" a call. Plus, it is being used to improve the interface to services such as voice mail, mobile Web services, telematics and automated attendants. Speech technology also will transform services, such as unified messaging, into easy-to-use mass market services. Adding the voice-activated interface to telephony services will be a stimulus to adoption and an important driver of increased customer satisfaction for years to come.

## **CHAPTER III**

### **Market Opportunities And Strategies**

#### **3.0 Telephony and Speech Technology Contexts**

The market for the application of speech technologies to the telephony arena holds much promise. As previously discussed, the technologies and their applications are now ready for mass consumption. Demand will come from both public telephony users and from corporate buyers. One of the segments with greatest potential is wireless.

The speech technology firms are already convinced that voice commands, not mobile phone keypads or mobile device keyboards, will become the main way wireless users will get data from their wireless devices. The new voice portal firms and voice ASPs have certainly embraced this same opinion and have acted on it. And now the wireless carriers are talking the same way and showing commitment through recent launches of new voice-activated dialing and portals. The optimists believe that most wireless data entry will be replaced by dependable and easy-to-use voice commands. Through next-generation phones and through network offerings, consumers will be able to use voice commands to make calls, access email and get information from the Internet.

On the broader level, information technology (IT) players are very optimistic about the impact of speech technologies generally on all forms of computing and communications. In their forecast article on emerging technologies, Computerworld lists speech technologies as one of the top five most important technologies for the next

decade. Wireless handhelds and speech recognition are viewed as two of the key technologies that will transform all aspects of information technology and communications in the next decade, the same way PCs did in the 1980s and the Internet did in the 1990s.

But the hype surrounding information demand on mobile devices has been robust. The wireless Web, mobile email, and mobile commerce are the basis for much of the drive toward upgrading to third generation (3G) wireless networks. But the demand must be real for the revenues to be forthcoming. This will be challenging in the coming years. So far the voice portal firms are starting small. But market tests and experience with market demand will determine the outcome.

The real demand for instant information must center on things that people cannot wait for until they get back home or the office. Messaging is one obvious answer. This includes getting voice-activated email and voice mail from anywhere. But it also means developing better ways of retrieving important alerts and transaction-related messages.

People need to be informed about the progress of certain transactions, such as a pending house closings. Or they need to be made aware of certain opportunities, such as a house for sale, a job opening, the next flight available, or where hotel rooms are available.

Application developers are focusing on these types of personalized and unique data delivery scenarios. For example, they envision a system where a travel agent would send the traveler that is in route an alert that explains that his or her flight has been delayed by two hours. And once they get that message they could give a voice command, call the booking agent and book a new flight. These solutions blend the emerging capabilities of speech recognition with wireless voice, SMS, EDA and location-based features.

### **3.1 Benefits of Speech Recognition**

The benefits of speech recognition are fairly obvious. Using speech recognition, systems:

- Eliminate confusing menus and system prompts that are difficult to remember;
- Let people interact more comfortably and naturally with automated systems;
- Provides automated 24 X 7 customer service; and
- Automates routine tasks in self-service implementations.

For the enterprise customers, call center managers can re-deploy and reduce the need for agents and they can cut call center costs while improving productivity. For end users, voice-activation can become the preferred way of interacting with telephony and data systems.

### **3.1.1 End User**

End users benefit from speech technology in a number of ways. One of the most straightforward is when they call into contact centers. In many systems, speech recognition is used in combination with IVR, but it still allows users to escape the limitations of burdensome voice menus. IVR menus can be both time-consuming and confusing when the menu choices get long and complex. The "barge-in" feature lets the user interrupt and override any current voice output with a new control command, eliminating the frustrating wait until the output has been completed.

Speech recognition as an input mechanism has come a long way since the early days which were dominated by speaker-dependent technologies, or limited-vocabulary, speaker-independent capabilities like speaking only the 0-9 digits or just "yes" or "no." With greater computer power, speech recognition technology can accommodate a lot of the variations found in "natural speech."

For the mobile user, voice command can be very useful as a hands-free way to control all communication functions, including voice and text messaging, as well as call control actions (e.g., voice dialing). It will also become useful in the future as an

alternative mode of delivery for time-critical alerts and responses sent to "on-the-go" recipients. We might need to coin a new phrase like "Short Information Service" (analogous to short message service) to describe the myriad ways that information delivery will benefit mobile users.

The mobile user has been touted as an ideal candidate for basic text-to-speech technology. It can be used to replace SMS sent for news alerts or the sending of email text messages via telephone. The conversion of text messages to speech is valuable to anyone who has a traditional voice-only telephone or where "eyes-free" retrieval is required (e.g., while driving a car). The increasing acceptance of Short Message Service (SMS) has already proven the need for delivery of short messages.

#### **3.1.1.1 Security**

One of the big concerns surrounding remote access to online data is security. Secure access to information along with personal identity protection could be handled by voice authentication. Enterprises are always concerned about protecting their databases while enabling mobile employees the access they need. Service providers need to be sure about the identity of their subscribers as they sign into various subscription services. Now that mobile devices have personalized both communications and information access, "identity theft" can become an even bigger problem beyond the hijacking of credit card numbers.

But bringing to bear the power of improved speech technology can simplify the sign-on and authentication processes for mobile users. More importantly, the unique convenience of a personalized voice print for all mobile contact and information access can provide greater protection against all forms of loss, including identity theft.

#### **3.1.2 Enterprise**

Enterprises are also reaping significant benefits from deploying speech technology. One of the most important is the provision of pleasant and easy-to-use interfaces for callers who call into the company or its call centers. This includes basic auto attendant systems, but also applies to the more advanced virtual personal assistants which serve customers, suppliers and employees. Benefits are well documented in applications for contact centers in a variety of industries, where evidence of higher levels of customer satisfaction drive interest in speech technologies.

Better customer service can be attained by supplementing or eliminating touchtone systems with speech recognition systems. When ease-of-use improves, customer satisfaction improves. Many users still give up when they get lost in IVR branching menus and "zero-out", that is hit 0 to get an operator. This defeats the purpose of automating contact centers.

The major speech technology vendors have case studies that document return on investment. Certainly companies can achieve cost reduction. The overhead of call centers is expensive, with unpredictable call volumes and the need for skilled agents. The downside of poorly designed systems is long on-hold times and limited service hours. Speech systems can save significant dollars. The basic ROI facts given by the industry include:

- Reduce costs by as much as 90 percent (comparing automated calls with operator-assisted calls);
- Cost of agent calls are from \$1 to \$15, while the cost of speech-enabled calls are from \$.10 to \$.85 per call;
- Yearly costs per speech port of \$3,500 versus yearly costs per employee of \$35,000; and
- 100 percent return on investment within a few months.

One example case study is Thrifty Car Rental. Thrifty reported saving \$1 million per year using speech technology. Using SpeechWorks software the automated system handles incoming customer requests for car rental quotes using speech recognition. Customers who call Thrifty's reservation number can obtain quotes through the

automated system and then are transferred to live agents to book the reservation. Thrifty receives four million calls per year and about 30 percent of those are for rate requests.

### **3.1.3 Carrier/Service Provider**

The key strategy among service providers relates to speech technology and wireless services. Voice-enabled services are viewed as a way to stimulate demand for wireless calling. By offering voice dialing, voice portals, and telematics, subscribers will use their mobile phones more. All the current voice portals, wireless Web sites, and telematics services make customers use their wireless minutes while using these services. Thus a customer's total minutes of usage will increase. But there are other possible strategies being addressed for speech technologies as well. The core strategies that are discussed by the stakeholders in this market are:

- Reduce churn;
- Stimulate usage;
- Effect differentiation;
- Create new service revenues; and
- Reduce risks.

#### **3.1.3.1 Reduce Churn**

According to studies funded by the wireless carriers, churn continues to be a major problem. Rates of 30 to 40 percent per year are not uncommon among the carriers. Compounding the problem is the fact that the higher income and higher usage customers are the ones more likely to switch. However, in reaction to this challenge, some companies have begun to take action to reduce churn. And one of the tools they are considering in the battle against churn is enabling their services with voice activation. By improving the ease of use with voice commands, subscribers will be less frustrated.

Voice-activated services will help attract customers and will help subscribers remain loyal. If voice dialing were offered, for example, customers would face the barrier of re-doing their phone directory if they switched to a new carrier.

The estimates of cost savings related to improved retention and less churn are significant. If one estimates that the cost of acquiring a new customer is \$300 to \$350, the total dollar savings are substantial. With rates of 30 percent or more per year, that equals millions of customers per year that must be replaced with new customers. One would add together the cost of acquiring new customers to replace the lost ones, plus the lost of extra revenues from the longer, established users.

In concert with their strategy to enhance services, carriers are launching data services, including messaging and access to the Web. And a widely discussed topic among carriers is the need for voice-activated interfaces to these services. The combination of useful information services and voice interfaces promises to reduce churn by creating "stickiness" with the consumer. Consumers will stay with their carriers longer. Carriers are now advocating the need for these value-added services as a major way to create customer loyalty. The basic rationale is customers that only use cell phones for voice calls, do not have any reason for loyalty to their cellular carrier. But those subscribing to enhanced services do.

Since it has the longest history, Japan's NTT DoCoMo is the prime example for success in enhanced services. Its churn rates dropped after the introduction of i-mode data services. DoCoMo now services over 15 million i-mode customers, who access a wide range of fun and useful messaging and Web-like services. In addition, NTT has found that i-mode customers generate more revenues per month. A DoCoMo customer adding i-mode generates significantly more revenue than a voice customer. And at the same time it helps in reducing churn. It reports that churn among i-mode customers has dropped to less than one-tenth of one percent per month.

For most carriers, price has been the primary way that they have differentiated themselves in the market. But price is the number-one reason customers churn to other carriers. Industry experts are beginning to see a parallel between the mobile phone



market and the long-distance telephone market, which once hinged its entire value proposition on price. This resulted in customers switching carriers whenever they were offered a better deal. Now that the price of cellular service has dropped and become so widely available, the same phenomenon has occurred. Customers can easily switch to get the best price. However, carrier marketers notice that customers do select on other variables as well. This is the same pattern observed among long distance companies. As prices dropped, customers began to expect low prices but also wanted more. Customers began to expect good prices in the long-distance company but wanted something more to enhance their basic service.

### **3.1.3.2 Stimulate Usage**

A key strategy for carriers is using voice-activated services as a means for stimulating demand for service. This can be accomplished in several ways. One way is to promote use of the stand-alone voice portals. A second way is to create and host their own voice portal services. A third is to link wireless services into the unified messaging strategy. And the fourth is to increase demand by selling into telematics applications for wireless connections in vehicles.

In recent wireless conferences, the issue of continuing decreases in average per customer revenues has been widely discussed. The average revenue per user or ARPU has fallen over most of the past decade. Wireless carriers have spent and will be spending new large sums of money to buy 3G spectrum to serve a subscriber base that is paying less and less per minute. In major strategy plans going forward, all the carriers need to stimulate usage to boost this decline. The strategies include implementing speech recognition-related applications. If the evidence collected from the initial efforts holds, these strategies will work.

### **3.1.3.3 Effect Differentiation**

Certainly speech technology will enable service provider differentiation. Providing easy-to-use voice interfaces will increase competitive differentiation and allow for network branding. One example on the enterprise side is Mitel. Mitel recently announced their Speak@Ease product portfolio to be offered in conjunction with their enterprise PBX and voice mail products. Mitel is committed to speech recognition as a competitive differentiator and expects to drive increased sales as a result of the advanced auto attendant capability and secure voice navigation of messages and other personal information.

#### **3.1.3.4 Create New Service Revenues**

A host of innovative service offerings were profiled in Chapter II above. These include the voice-driven personal assistants made by Webley, Call Sciences, General Magic, and @Motion. They are expected to help telcos capture subscription revenue by offering multiple, voice-driven services in one package (call screening, unified messaging, conferencing, address book dialing, calendaring, content, and more).

Innovative service offerings are also emerging with the use of speech technologies with directory assistance. One example is the firm Call FYI, which is experimenting with offering yellow page services over the phone.

#### **3.1.3.5 Reduce Risks**

Using speech technology has become a necessity in certain markets to insure regulatory compliance. Manual dialing while driving will soon to be illegal in over 10 countries throughout the world. To let consumers safely use their phone while driving, voice-activated dialing packages are nearing official rollout in Australia and the US. These offer easy number-dialing and address book capability.

Companies are reducing their investment risk by not hiring costly operators and by automating incoming calls instead. They are reducing variable costs-per-call because speech recognition lets consumers interact with the system as long as they would like.

## **3.2 Value Chain**

There are a host of different types of companies that provide speech solutions across the telephony value chain. They range from niche players that provide voice-activated email systems to system integrators that implement suites of software products that turn speech technology into powerful business solutions. The telephony speech market place is really an intersection of literally dozens of applications, markets and players

The goal of this section is to sort out the extremely complex set of players in the arena of speech technology adoption and use. This is important to do for a number of reasons. One is that it sets the stage for assessing opportunity and determining which revenue streams to portray. Secondly, it helps determine who are the key providers of the core technology to this marketplace. And third it is used to determine which firms are affecting the various segments of the telephony world, like call centers, messaging services and directory assistance.

Part one of this explanation will use the value chain model to describe the types of players and distinguish between them. Part two will profile the major speech technology vendors – those firms that license speech engines to the rest of the value chain. And part three will cover the specialized firms that focus on specific segments and application areas, such as unified messaging or directory assistance. This part will profile some of the players that The PELORUS Group considers to be the best-of-breed firms in each of the important telephony application areas.

### **3.2.1 Value Chain Analysis**

The value-chain notion has a different focus, and a larger scope than the more traditional idea of supply chain. Value chain analysis looks at every step of a market delivery from raw materials to the eventual end-user and even to details such as disposing of the packaging after use. The goal in value chain analysis is to understand how to deliver maximum value to the end user for the least possible total cost. This level of analysis is too comprehensive for this report, but the issues must still be addressed through a discussion of three viewpoints.

The various types of speech technology, such as speech recognition or text-to-speech, are really often used simply as components in a more sophisticated set of application software that ultimately enables functions such as automated call centers, for example. Speech application software is itself part of larger automation or enhanced telephony architectures that include hardware and software elements that merge telephony, speech and enterprise elements.

### **3.2.1.1 Service or End User Viewpoint**

It is important to realize that the application of speech technology to the telephony arena consists of a complicated set of intersections. In this report, the major emphasis is the market view, yet the major focus is on a set of technologies. As the development, adoption, and implementation of speech technologies evolves clearer boundaries may come about, but in the near term, much remains to be determined.

What remains clear, however, is that speech affects all aspects of product and service offerings. Speech technology is a clear winner in its ability to change the user interface. As such, it is simply a service enhancement. This means it is difficult to quantify in its impact and revenue streams. Speech technology can also be used in combination with other interfaces and services to come up with unique new versions of services. For example, when mobile handhelds are used to find and deliver email to the end user by voice command. In a few instances voice-activation is used in creating actual new

services, such as voice-activated dialing. Without the speech interface, you do not have the service.

In Figure 3.1, varetis tries to develop a model for portraying the way in which speech intersects with the telephony service model. The top row shows the various access channels that are offered by telephony service providers. Users can access services in a number of ways, including by WAP, SMS, or voice automation. The services provide access to various end user applications, such as White Pages, as seen in row 2. Middleware and application firms provide the support in row 3, often supported by third party components. The bottom level or back-end provides the linkages out to supporting systems.

### **3.2.1.2 Company Viewpoint**

The viewpoint from the perspective of the speech technology or applications provider is also important. Companies cannot implement speech solutions alone. In fact they often point to important partnerships as a way to prove their success. For example, IBM's WebSphere product offers a range of solutions for firms running Internet applications. This includes voice-activated versions. Figure 3.2 shows their view of how the addition of speech technology impacts the value chain.

The IBM WebSphere Voice Value Chain runs from the basic telephony platform on the far left to the end user consumer or business customer on the right. There are 10 pieces of the value chain running from left to right, plus additional partnerships below some of them. IBM provides middleware that sits on top of the existing telephony platform vendor products. The WebSphere Voice Response and the WebSphere Voice Server must translate traffic coming from the telephony networks so it can be utilized in speech applications. IBM's speech technology software is embedded into their middleware that creates voice-activated applications. But these only affect the functional level. The value chain moves into the solutions arena where another set of players must

contribute the voice applications, along with the content sets from its partners, such as ISx directory assistance offerings.

### **3.2.1.3 Industry Viewpoint**

The final perspective is from the telephony industry as a whole. Figure 3.3 is used as an illustration of how to portray the complex issue of speech stakeholders. Although this model is taken from the viewpoint of Internet services, the same model holds for telephony services. As the diagram shows, there are five primary sectors that take part in the value chain, represented by the five segments of the circle. The first three -- making up the right-hand side of the circle -- have direct contact with the user, while the remaining two -- making up the left-hand side - stay in the background, as operators of the Internet (or telephony) infrastructure.

Application providers create the software and applications from which solutions are assembled. This category encompasses the software developers and independent software vendors (ISVs), whose products ASPs and service integrators deliver. The segment also includes providers who host and deliver their own applications as online services.

Service integrators are the providers that end-user businesses recognize as application service providers (ASPs). They bring together services for delivery as complete, managed solutions. They range in sophistication from enterprise ASPs to commercial voice portals. This category also includes more traditional types of professional services company, such as service aggregators and end-user or full-service providers.

Access providers take care of the 'last mile' connection that allows users to access the network. This category includes telecom providers and Internet service providers, many of whom bundle access along with a basic suite of hosted applications. It also includes wireless ASPs, who deliver Web content and applications to mobile users.

Infrastructure operators look after the physical backend elements of the network. They include the telecom providers who act as Internet backbone carriers, as well as

collocation and hosting providers, who manage Internet data center facilities. Another group in this segment provide utility-style computing resources, such as pay-as-you-go storage services.

Infrastructure service providers make up the software and services layer of the Internet computing infrastructure. They include application and ASP infrastructure providers (AIPs), who operate hosting centers that are specially equipped for application hosting. AIPs often work with infrastructure ASPs, who deliver specific elements of the infrastructure, such as billing and metering, directory services, or payment processing. Management service providers, who specialize in remote management of IT systems, are another fast-growing group in this category.

A final group has been left out of this picture, but they remain an important component: research centers and universities. For speech technology firms, the main benefit of EU research programs is the knowledge and experience gained from working with partners from along the value chain. On the one hand, universities and research institutes bring fundamental knowledge and new theoretical findings; while application providers and users provide valuable feedback on market requirements.

### **3.2.2 Voice (Speech) ASPs**

Voice application service providers (ASPs) provide pre-packaged voice applications and products -- everything from call center solutions to PBXs. These products are usually customizable by the vendor or one of their systems integration partners. Some of the application partners offer OEM solutions that include speech software bundled with their own products. These bundled solutions are lightly configurable and not customizable. This category includes software developer firms and independent software vendors whose products are used by VARs or system integrators to deliver product and services.

ASPs are expected to be among the high tech industry's fastest growing sector in the coming years. ASPs can be compared to Internet Service Providers (ISPs) in how they operate. ISPs are providers of network-based applications such as email, Web site

hosting and design, chat and bulletin boards. They build and operate applications for the network carriers or corporate enterprises.

Companies that want to provide public access to their information over phones will not have to invest in telephony infrastructure. Just as they do not have to buy Web servers to launch their Web sites, but rather buy services from an ISP. Now they will be able to buy voice sites from voice service bureaus. These can be thought of as voice application service providers or voice ASPs. Some of the voice ASPs may actually emerge from existing ISPs, who see a natural progression from Web-only to multimedia hosting opportunities. And Voice ASPs are also emerging from IT service start-ups who see themselves as data storage and presentation ASPs.

The rapid emergence of these service firms is a driver for rapid expansion of speech recognition in the wireless markets. Application service providers are enthusiastic about adoption of new speech technologies in the wireless markets. ASPs create, deploy, run and maintain applications for carriers. ASPs are active in deploying:

- Unified messaging solutions;
- Wireless Web access;
- Voice portals; and
- Corporate network and mobile worker solutions.

Several of the voice portal firms are evolving into voice ASPs. BeVocal has both a direct phone service and is an ASP for Qwest. ASPs can build private-label portal services for the carriers. Instead of investing in their own speech recognition technology, carriers use ASPs to manage and run software from their servers to save on costs and service personnel. Speech recognition is a complex technology and requires a large amount of maintenance to make sure that it operates optimally. ASPs offer applications that can be used across a number of vertical industries. ASPs sometime offer their services on a per call basis depending on the types of calls. For example, Voci (Campbell, California) charges from \$.50 to \$1.25 per call. ASPs are also aiming their consulting at Web portal companies too, such as Yahoo!, Excite and Netscape.



General Magic (Sunnyvale, California) offers a voice ASP service based on their VoiceXML 1.0 MagicTalk platform. General Magic is known for their personal assistant service, called Portico. They have incorporated speech software from Nuance, IBM and SpeechWorks. The platform is built on IBM's WebSphere Voice Server and IBM's DirectTalk Voice system. General Magic's solution will also be used in GM's OnStar Virtual Advisor service.

Price Interactive (Reston, Virginia) is the company that has created MapQuest's voice-enabled site. They have also served as an ASP to the IVR service bureau world. They are SpeechWorks partners in IVR and have partnered with Lernout & Hauspie in their RealSpeak text-to-speech software. Price Interactive offers its SpeechPort service to major Web site owners. The SpeechPort has carrier-grade facilities in St. Louis, Missouri and Fairfax, Virginia, plus a new site in Denver, Colorado. In August 2000, they bought the e-business firm, FusionTec.

One application developer, VoiceGenie, has already released VoicePortal+, which is a voice portal tool that could be sold to voice portal service firms, such as Yahoo!, Tellme, etc. VoicePortal+ is a content aggregator that offers access by wireline, wireless or cable phone connections.

This is only a sampling of the efforts by various firms to implement speech recognition solutions. Many different firms are players in the value chain, including players that can be classified as network providers, device and equipment manufacturers, ISPs, middleware firms, software firms, location service providers, plus content, portal and e-commerce firms.

### **3.2.3 System Integrators**

Value added resellers (VARs) or system integrators firms develop and customize speech-based applications and integrate them into computing and operations networks for customers. Many are also capable of integrating component technologies into a deployment platform for customers. Some system integrators work with application

providers and/or other speech technology partners. VARs typically prime the project and deliver the complete voice-driven solution, including after-sales support. Some of the VARs active in speech deployments are: General Magic, Gold Systems, Edify, Price Interactive, VoiceGenie, Preferred Voice, Phone.com, NetbyTel, Rhetorical Systems, Verascape, and West Corporation.

### **3.3 Key Players and Strategies**

Using the value chain model, we have identified two core groups of players to focus on, namely those that provide the speech technology and those offering applications. The first set of profiles below relates to those set of players that has developed and sell software, consisting of the core software engines that drive the speech technologies -- automatic speech recognition, text-to-speech, speaker verification, and natural language understanding. The second set of profiles is a group of key players that The PELORUS Group has identified as leaders in their application areas. Selected from among the ten most important applications markets (e.g. voice dialing, directory assistance, auto attendant, and telematics), these companies are considered "best-of-breed."

#### **3.3.1 Speech Technology Vendors**

This section covers the key speech technology vendors and contains profiles on twenty companies ranging from small startups to large multinationals. Nuance and SpeechWorks dominate the telephony speech technology landscape in the United States, which is still the largest market for speech applications. Both Nuance and SpeechWorks represent firms whose mission is speech technology. As an example of the wide set of products that comes out of one of these firms, Nuance has agreed to have us reprint a top-level description of their products. Figure 3.4 shows their product set.

Speech technology is also offered through one of the divisions of such major corporations as: AT&T, IBM, Lucent, Microsoft, Motorola, and Philips. Figure 3.5 shows

the range of products from Philips as an example of the kinds of products coming from these firms.

Included among these profiles are two French firms, Telisma and Elan, and the UK firms of SRC and Vocalis Group. These firms are on the whole all doing well, except for the major speech technology pioneer Lernout & Hauspie (L&H). The Belgian-based firm has gone through bankruptcy proceedings and its assets have been sold. ScanSoft bought the major speech technology assets of L&H. Each profile contains our perspective about the company's position in the market and about the latest speech technology offerings related to telephony applications. Each profile also includes short summaries of key developments that occurred during the past year, 2001.

### **3.3.1.1 AT&T Labs**

Established in 1996 after the split of AT&T and Lucent, AT&T Labs, with headquarters in Florham Park NJ, has 6,000 employees. The group that works on speech technologies numbers around 500. Recently AT&T Labs took the bold move of selling its technology directly into the marketplace, rather than remaining a R&D unit only. AT&T Labs also has a new president. In August 2001, AT&T gave its top technology post, President of AT&T Labs, to Hossein Eslambolchi, a veteran engineer from the Bell System. Eslambolchi took the helm from outgoing CTO Dave Nagel, who left AT&T to head a new operating systems division at Palm. At his last post, Eslambolchi helped dramatically improve the network at AT&T's cable ISP affiliate, [Excite@Home](#).

AT&T Labs' core product, its Natural Voices Text-to-Speech Engine, was launched in July 2001. It is positioned by AT&T as the most realistic, human-sounding of the synthetic speech systems available on the market today and according to attendees at the SpeechTEK 20001 show in September, many in the industry agree. The Labs' technology was even used recently in a Warner Brothers movie, called Red Planet.

The speech engine comes in three distinct packages targeted at different size organizations and deployments. The Server edition is designed for use by large

businesses serving the needs of many users across an enterprise network. The Server-Lite edition is aimed at application developers and small businesses serving as many simultaneous users as the hardware allows. The Desktop edition is intended for developers of end user desktop applications and individual users. All editions of the TTS Engine have multilingual capabilities and include a male and female voice in U.S. English. An optional female Spanish voice was made available at the end of 2001.

AT&T Labs also offers a customization module that gives users of AT&T Labs' Natural Voices TTS Engine a way have their own unique voice. This means that changes can be made to the voice quality and sound so it meshes with existing corporate image and branding. AT&T Labs calls this capability the Natural Voices icon becomes once customized, the voice becomes the sole property of the customer who orders that specific voice. Otherwise, businesses can use the male and female voices that AT&T Labs calls voice fonts that are offered with the Natural Voices software.

The AT&T Labs Natural Voices product line was just launched during the second half of 2001. AT&T Labs TTS technology is licensed and sold by many companies in the industry. The three Natural Voices products and their details are:

1. Server Edition -- AT&T Labs' Natural Voices TTS Engine Server edition serves large businesses working on a client/server architecture making it useful to all types of applications on the network. Plus it supports many users accessing different TTS-enabled applications simultaneously. The Server edition supports the most popular computer and development platforms including Windows NT, 98, 2000, ME, XP; Linux 6.1, 6.2, 7.2; Solaris/SPARC 2.7, 2.8; HPUX 11.0, and AIX 4.3.
2. Server-Lite -- The Server-Lite configuration is ideal for a small business environment where there might be several people who need simultaneous access to the same application. It supports the most popular computer platforms including Windows NT, 98, 2000, ME, XP; Linux 6.1, 6.2, 7.2; Solaris/SPARC 2.7, 2.8; HPUX 11.0, and AIX 4.3.
3. Desktop – The Desktop edition lets one application speak at a time and is targeted at the low end of the market: individuals and developers who wish to add

TTS capabilities to their desktop applications. The Desktop Edition supports Windows NT, 98, 2000, ME, XP and Linux 6.1, 6.2, 7.2

### Key Developments

- One Voice Technologies, a developer of 4<sup>th</sup> Generation voice solutions for the wireless and telematics firms, confirmed that its MobileVoice Platform would be the first messaging platform to integrate with AT&T Labs' Natural Voices text-to-speech engine. The goal, according to One Voice officials, is for a more flexible and natural sounding voice than has been possible before. This should help pave the way for faster user acceptance and higher market penetration of its products.
- AT&T Labs signed an agreement with Centerpost Corporation to use its Natural Voices Text-to-Speech technology to deliver automated voice alerts to travel industry customers. The use of automated voice alerts in travel situations is one of the key applications for Natural Voices, since calls in this environment are typically urgent in nature.
- VoiceGenie, an important ASP in the wireless and wireline spaces, announced that it would integrate the AT&T Labs TTS product into its VoiceXML gateway. This will be used for creation of voice portals and other m-commerce applications.
- The TTS engine has been made compliant with Microsoft's Speech Application Programming Interface (SAPI 5.1) and works immediately with any SAPI 5.1 application on the desktop. The software development kit for all three editions enables developers to create SAPI-compliant applications or use the C++ Application Programming Interfaces to create TTS-enabled applications.

#### **3.3.1.2 BBN Technologies (Verizon)**

Cambridge, Massachusetts-based BBN's roots go back a few decades and can be considered a pioneer in the Internet and networking areas. Known as Bolt Beranek and

Newman (BBN), the company was founded in 1948, and offered high-speed dial-up access, systems development, network security, and Web hosting services. BBN has been responsible for a number of innovations. It designed the acoustics for the UN General Assembly Hall. It designed, implemented and operated the ARPANET, the forerunner of today's Internet; and pioneered the first packet switch, the first router, and the first person-to-person network e-mail. BBN also created an icon for the digital age—the @ sign. The BBN Systems and Technologies unit was purchased in 1997 by GTE. Today, BBN is part of Verizon Communications and employs approximately 700 people in 10 locations across the U.S. Revenues for fiscal year 2000 were \$118 million.

BBN's Speech & Language Processing Department continues its development of speech recognition and natural language understanding technologies. They have developed speech solutions in English, Spanish, Chinese, and Arabic. BBN expertise includes: speech recognition, audio indexing, speaker verification, dialog processing, language understanding, data extraction from text, name finding, topic spotting and optical character recognition.

#### Key Developments

- BBN and SAIL LABS Technology AG, of Vienna, Austria, signed a licensing agreement that gives SAIL LABS Technology rights to use and modify current and future BBN real-time speech and language technologies. SAIL LABS Technology is a major European supplier of natural language understanding products, such as speech recognition, dialog management, and media indexing.
- BBN Technologies and Rockwell Electronic Commerce announced joint marketing of a new review and analysis process product called Explorer IVR Automation Assessment. This service will help large call centers improve their IVR systems and design new ways to route callers to appropriate representatives more quickly and efficiently. The product will be sold to businesses that have deployed interactive voice response (IVR) systems in large customer-contact centers. Once a business signs on for Explorer IVR Automation Assessment, BBN and Rockwell

Electronic Commerce experts will evaluate the usability and potential cost savings of IVR systems.

- BBN launched a new advanced language processing technology, called BBN Call Director. It is designed to replace touchtone menus in call centers and differs from other call-routing solutions because it recognizes, understands and responds to a caller's spontaneous speech, unlike other systems that depend on fixed, limited grammars.

### **3.3.1.3 Elan Informatique**

On November 8, 2001, Elan Informatique won a decisive round in its lawsuit against Lernout & Hauspie, opening the way for a return to Elan's total independence. Elan Informatique, a provider of text-to-speech technology, won its case against its shareholder L&H for a return of its own shares of stock. At the present time, the L&H group is in liquidation, so Elan's shareholders claimed restitution of their shares in compensation for the demise of the business. This recent court judgment enables Elan to continue business as usual and pursue on-going research and development projects with other industry players.

Currently, Elan Informatique, based in Toulouse, France, has about 300 customers, with 65 percent of its customers in Europe, 15 percent in North America and 20 percent in South America. Its core products are Speech CUBE and NetWorks CUBE. Speech CUBE is a TTS software component for enterprise and telco applications. Speech CUBE is multilingual and multichannel, transforming any IT generated text into speech. It has capacity for handling 10 languages and 17 different voices. The languages are: American English, British English, German, French, Castilian Spanish, Italian, Polish, Russian, Brazilian Portuguese and Latin American Spanish. It can read proper names, numbers or currencies. It can be used to convert email, HTML or RTF content. A single machine can handle 400 ports. It utilizes Unisys software. NetWorks CUBE is designed for large-scale

deployments in telcos such as those for voice portals, directories or SMS to voice conversions.

#### Key Developments

- Elan will be showcasing its new generation version of its TTS technology early in 2002, with an offer aimed at challenging the highest existing TTS quality today.
- With its Latin American Spanish TTS, Elan continues its long-term strategy of developing languages that respond to clearly defined business needs and bringing them to market when and where they are most useful. Elan Informatique had already planned to deploy over 3000 TTS ports in South American in 2001.
- Elan Informatique finalized a 4-year license agreement with SIEMENS VDO for its multilingual text-to-speech technology. The application of the Text-To-Speech (TTS) in the modular driver information system of SIEMENS VDO includes entertainment, communication and navigation functions. The TTS will be used to read SMS (short messages), e-mails, traffic messages and WML content.

#### **3.3.1.4 Fonix**

Founded in 1994, Fonix Corporation is based in Salt Lake City, Utah and employs 115 people. The company develops text-to-speech, handwriting recognition, and automated speech recognition software that are integrated into a variety of products for commercial and industrial applications. Fonix's software is targeted at two core markets: embedded systems applications (electronic devices and consumer products) and server-based applications for Internet and telephony voice-activated functions, including call center automation. After investing heavily over the last couple years in research and development, Fonix hopes to begin to see significant generation of revenues through royalties, licensing agreements, and product sales. As part of its international expansion, in 2001 Fonix opened an office in South Korea.



The key Fonix product is Fonix Accelerated Applications Solutions Technology (FAAST) Text-to-Speech. With FAAST Embedded 1.1 and FAAST Windows CE 1.1, developers can integrate Fonix Automatic Speech Recognition (ASR) and Fonix Text-To-Speech (TTS) into embedded applications including cell phones, Personal Digital Assistants (PDAs), wireless devices, appliances, and automotive applications. Fonix neural network technology allows accurate speech recognition in noisy environments and does not require the user to train his or her voice to the system. Customers can also create command-and-control applications that respond to continuous speech, enabling users to speak naturally to the device. Fonix customized ASR allows the user to modify the application in any language with his or her voice.

The FAAST framework, including Fonix TTS and ASR engines, is marketed primarily to solution integrators, including OEMs and VARs, who bring together various applications and products to provide a complete solution to their end-user customers. Fonix has entered into agreements with several key solution providers such as Nortel Networks, Nuance Communications, Envoy, and Motorola Mobile Internet Exchange for one or both of these engines. Fonix continues to focus on product enhancement and application delivery in the following server-based markets:

a) Interactive voice response (IVR) - Telephony applications for corporate call centers, short messaging services (SMS) and information retrieval from server databases. FAAST facilitates upgrades and shortens development time required for new systems.

b) Internet Voice Portals - Speech-enabled access to the Internet and Web site navigation. Fonix is developing and anticipates marketing products for voice portals to the Internet that may be purchased for use by portal companies, website and content providers, Internet service providers and browsers.

c) Website readers - Because over 70% of all web content is text, demand for products allowing web content to be read to the user is quickly growing. Fonix has developed a web page reader that can be added to websites. This voice solution is

targeted toward the largest web content providers worldwide in media, government, business and other industries.

d) Network systems command and control and email reader - Fonix provides TTS which can be seamlessly integrated into network software to speech-enable software applications.

### Key Developments

- Fonix made a number of agreements poised at the telematics market. They reached a distribution agreement with Motorola Semiconductor Products Sector (SPS) for advanced speech solutions delivered on Motorola's mobileGT platform, which is Motorola's telematics solutions division. This agreement allows Motorola SPS to market and sell Fonix Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) applications and solutions to Motorola customers worldwide. The agreement is expected to increase the current channel distribution of Fonix voice-enabled automotive solutions and applications worldwide. Fonix expertise in speech enabling in-vehicle applications and Motorola's extensive global sales force, are expected to capture jointly significant market share for speech solutions in the automotive space."
- Fonix and InfoSpace are integrating speech technologies to offer solutions to wireless carriers. Fonix FFAST will be integrated into InfoSpace Speech Platform and TTS server.
- A joint effort between Fonix and Panasonic has been released. The Panasonic KX-TG2680N telephone, featuring the new Talking Caller ID solution, debuted at the CES Show in Las Vegas in January 2001. Sales exceeded Panasonic's expectations in first half 2001 by nearly 100%. Panasonic and Fonix are continuing to develop additional low-cost solutions for future product lines of cordless phones and other consumer electronics.

### **3.3.1.5 IBM Voice Systems**

IBM's Voice Systems (West Palm Beach, Florida) is responsible for a number of speech technologies, including speech recognition for telematics, the Internet and telephony markets. IBM's speech technology middleware portfolio, called the WebSphere Voice Family, includes five components: IBM WebSphere Voice Server, IBM WebSphere Voice Toolkit, IBM WebSphere Voice Response with DirectTalk Technology, IBM Message Center, and IBM WebSphere Translation Server. The speech technology server, WebSphere Voice Server, was launched at the end of 2000 and is IBM's core speech technology and e-business infrastructure software. It resides on either their IVR platform or VoIP gateway. In fact, IBM was the first to offer speech technology for Voice over IP systems.

In October 2001, IBM launched the latest version of its Voice Server technology. IBM WebSphere Voice Server 2.0 includes four elements: speech recognition engine, TTS engine, voice application development tools and a telephony platform connector. New features include concatenative TTS that sounds closer to a real person, and support for more languages: Chinese (Mandarin), Italian, Japanese and Spanish, along with existing support for English (U.S., U.K. and Canadian), French and German. WebSphere Voice Server now supports Intel's popular telephony platform. The product was made available in November 2001 and costs \$15,000 per processor plus \$3,000 per language.

IBM also has re-introduced its IVR product, called WebSphere Voice Response with DirectTalk Technology, which is the latest version of its IVR software that accepts both voice and touchtone keypad response. Formerly known just as DirectTalk, the product can now handle higher call loads and has been brought under the WebSphere platform. WebSphere Voice Response with DirectTalk Technology supports both VoiceXML and Java standards. This product works together with WebSphere Voice Server for combined speech recognition and IVR applications, or by it to automate a range of call-center related services.

One example of IBM innovation is a call center customer is the financial services firm, T. Rowe Price. Participants in T. Rowe Price retirement plans who call into their customer service number get an automated phone attendant powered by IBM's latest voice recognition products. The system allows 401(k) participants to check fund and account balances, request statements and conduct other business by simply speaking their requests. In fact, in this implementation the natural language understanding feature of IBM's WebSphere Voice Server was used.

During the last two years, IBM has made progress in acquiring telematics customers and partners. Its embedded speech products will be making their way into new in-vehicle applications using voice control. Current projects include:

- Johnson Controls, a leading automotive and telematics supplier, has selected IBM as its speech software provider for all of its upcoming telematics offerings to the auto industry. Its first implementation is a voice-enabled mobile communications system for the Chrysler Group;
- Intel is working with IBM's Embedded ViaVoice recognition product to support its strategy to deliver a non-proprietary, standards-based in-car platform based on Intel StrongARM and upcoming Intel XScale processors;
- IBM's Embedded ViaVoice recognition will also be used on Hitachi's SH3 and SH4 processor platforms (called HARP) for automotive applications. Hitachi is a leader in the embedded solutions telematics market. Its systems will support over 10 languages for in-vehicle speech applications.

### Key Developments

- IBM claimed a number of new customers in 2001, in the areas of telematics (mentioned above), PDAs (Legend Computers and Compaq); contact centers (Maritz and Plaza Associates) and voice Web (Xora);
- IBM also supports the mobile telephony world through its IBM Embedded ViaVoice Mobility Suite. IBM technology has found its way into all kinds of mobile devices.

This year, Legend Computers, the largest computer manufacturer in China, chose IBM for voice-enabled PIM solutions on its mobile devices and Compaq chose IBM for voice enabling its iPaq Pocket PC;

- Alcatel and Genesys Telecommunications Laboratories, Inc., a wholly-owned subsidiary of Alcatel, announced in 2001 plans to purchase IBM's CallPath assets. CallPath is computer telephony software that integrates voice and data communications in contact centers. Genesys intends to continue support for the approximately 400 CallPath customers, while IBM remains the first point of contact for service under existing agreements. IBM and Genesys will also dedicate resources to offer current CallPath customers new contact center capabilities through the Genesys Contact Center Offering, which offers contact center solutions that use IBM products for unassisted service and Genesys products for assisted service.

#### **3.3.1.6 InfoSpace**

InfoSpace is a global provider of Internet infrastructure services and has three lines of business: merchant services, wireline and wireless. InfoSpace Speech Solutions group is part of wireless services unit. InfoSpace powers private labeled services to the leading wireless carriers worldwide, reaching 80% of US wireless users through relationships with Verizon, AT&T, SBC and ALLTEL and device and equipment manufacturers such as Nokia and Ericsson.

InfoSpace has expanded into speech technology through its acquisition of Montreal-based Locus Dialogue Inc., a developer of speech recognition-enabled applications, in January 2001. This acquisition propels InfoSpace into the speech technology business and allows it to speech-enable a wide variety of existing wireless applications, such as personalized information managers (PIMs), location-based services and mobile commerce services. In addition, this will allow InfoSpace to develop other unique mobile offerings,

such as speech-activated dialing. With the purchase, InfoSpace bought Locus Dialogue's core speech technologies and speech-enabled applications, which have been in development for more than 14 years. This includes automatic speech recognition, natural language understanding, speaker verification, and text-to-speech server technology. A part of the former Locus Dialogue continues as LocusDialog (see profile below). The Liaison product line from LocusDialog is a speech-enabled call routing system sold to enterprises.

InfoSpace is a big believer in the necessity of multimodal capabilities for the user, so its solutions include speech-enabled, SMS, WAP and Web presentations of information. InfoSpace Speech Solutions is able to deliver Internet content and directory information to desktops and mobile devices using any one of these alternative means. InfoSpace also adheres to a strategy of open platforms. Its speech platform can host any ASR, and is open to any vXML gateway or third party applications. InfoSpace already powers significant DA services via its Web services. But recently InfoSpace has made expanded its play in the directory assistance marketplace with a new partnership with the enterprise DA powerhouse, Phonetic Systems (see below).

### Key Developments

- Late in 2001, Phonetic Systems and InfoSpace established a joint marketing agreement to provide non-operator assisted, speech-enabled directory service applications for wireless carriers and other service providers. This deal will combine Phonetic Systems' ability to voice-enable large directories with InfoSpace's wireless and Internet software and application services. Phonetic Systems' patented Voice Search Engine will be combined with the InfoSpace Speech Platform and speech-enabled directory service applications. They finished the technical collaboration phase and did a trial in the Seattle area. Now they are out promoting the solution to wireless carriers.

- Fonix Corporation, a provider of natural-user interface technology and voice solutions, completed an agreement with InfoSpace, Inc. to integrate the Fonix Accelerated Application Solutions Technology (FAAST) Text-To-Speech (TTS) engine into the InfoSpace Speech Platform and TTS server. InfoSpace provides an open standards-based, VoiceXML speech platform and applications that enable the delivery and development of Internet-based consumer and commerce services accessible from any phone. By incorporating a high-quality TTS engines, such as FAAST TTS, into its speech synthesis TTS server, InfoSpace will be able to provide partners, developers and designers an opportunity to select the technologies that are cost-effective and most appropriate for a specific application.
- In 2001, InfoSpace also signed an agreement for joint R&D with Bell Mobility Investments (BMI), a Canadian corporate venture company created by BCE and Bell Mobility. The R&D activities are related to a VoiceXML-based voice activated dialing (VAD) application. Using a speaker independent speech recognition technology, the planned service will support simultaneous recognition of both English and French languages. The InfoSpace VAD solution will be robust, since it can be linked to an individual's personal information manager (i.e., calendar, address book and to-do list).

### **3.3.1.7 Lernout & Hauspie**

As of October 2001, Lernout & Hauspie (L&H) owned 154 technology patents. Based in Belgium, L&H has been a major, long-term purveyor of speech and language technologies. However, L&H is in bankruptcy proceedings and a determination of which assets will be sold off or retained has yet to be completely determined. Lernout & Hauspie's management has been charged with overstating its operating results and inflating its stock price for personal benefit. The findings led to bankruptcy filings by the US-based branch in Burlington, Massachusetts and its headquarters in Belgium. L&H has

said the most of its businesses will be sold, including: Mendez translation, Dictaphone Healthcare Solutions, Dictaphone Transcription Group, and Kurzweil Education.

On December 11, 2001 the U.S. Bankruptcy Court for the District of Delaware awarded most of the operating and technology assets of the Speech and Language Technologies division of Lernout & Hauspie to ScanSoft, Inc. of Peabody, Massachusetts. With the purchase ScanSoft obtains 200 L&H employees, and the critical text-to-speech and speech recognition technology. The Speech & Language Technologies group includes the Dragon products, which had constituted close to one-half of L&H yearly sales.

Despite these difficulties, L&H kept moving ahead during 2001. In 2001, it shipped a number of new products including a multilingual TTS engine and Dragon Naturally Speaking, version 6.0. Sales of its consumer speech products are sold mostly in the US and Europe. The breakdown is: 59% (US); 29% (Western Europe); 6% (Asia); 6% (Japan); 3% (ROW). In addition, L&H PDSay is soon to be released, which enables PDAs to use voice control for contact lookup, voice notification, and text-to-speech translation of emails. For Telcos, L&H has two core products: RealSpeak and Automatic Speech Recognition.

1) Text-to-Speech Solutions -- called L&H RealSpeak, it offers speech synthesis technology and toolkits that convert any computer-readable text into intelligible, human-sounding synthetic speech. L&H has over 300 customers of RealSpeak worldwide. They are available on a range of platforms including host-based (Windows or UNIX) and board-based (Antares). RealSpeak is available in some 19 languages, including: US English, UK English, German, Spanish, French, Italian, Dutch, Swedish, Norwegian, Korean, Japanese, Mandarin, Cantonese, and Portuguese. RealSpeak includes an email plug-in for developers who want to add email reading functionality to their applications. RealSpeak can be used in large-scale network systems, such as voice portals and mobile portals. But it is also applied in email, personal assistants, talking Web pages, unified messaging and CRM applications. In September 2001, L&H added a new male and female voice, called David and Deborah, to its TTS software to expand the American English version.



2) Automatic Speech Recognition Solutions – called ASR 1600, it is a consumer speech recognition engine that is designed for use in handheld computers and high-end phones. Designed to withstand noisy environments, it is used in PDAs and automotive applications. The ASR product is offered in 14 languages.

L&H focuses on embedded systems, but offers a variety of speech technology products. Its technology has been sold widely. Its voice portal technology has been sold to the major portal firms BeVocal, Tellme, and Internet Speech. L&H sold its speech technologies for unified messaging and email reading solutions to Cisco, Lucent, Comverse and Nortel. L&H has over 700 contracts with other manufacturers and service providers to integrate its technology in their offerings. It has licensed its technology to a long list of major firms, including the telecom and IT companies: Alcatel, British Telecom, Lucent, Intel, Deutsche Telekom, GTE, Microsoft, Motorola, NEC, Nortel, Periphonics, AOL and Unisys.

### Key Developments

- EPOS (Auburn, AL) announced a new partnership with L&H. EPOS is a provider of automated self-service solutions for call centers and has over 3,000 installations worldwide. Using IVR and speech technology, it facilitates communications via phone, Web, wireless and PDA devices. EPOS will add L&H RealSpeak to its FirstLine Encore communications platform.
- L&H's Speech and Language Technologies Division decided to make its RealSpeak Compact Software Development Kit (SDK) available to developers. This should make its TTS technology available for a broad range of mobile devices, such as PDAs, mobile terminals, and handheld computers. The SDK supports Microsoft Windows CE or Linux operating systems.
- Mid-2001, L&H Automotive Solutions Group announced the Distributed Speech Recognition (DSR) technology for the automotive market. DSR enhances the performance of mobile communications devices in vehicles by

adding voice command functionality. The telematics system takes spoken commands and transmits them as digital data over a wireless link to a network server. Processing is completed and the requested information, such as email messages, is returned to the driver. L&H claims that it has unique noise-robust data processing modules that ensure a high level of accuracy in recognizing voice commands for driving environments.

### **3.3.1.8 LocusDialog**

LocusDialog (Montreal, Canada), which was once known as Locus Dialogue, has been renamed and reestablished as an independent, privately owned company. When Locus Dialogue was acquired by InfoSpace in January of 2001, InfoSpace bought the core speech technologies to form InfoSpace Speech Solutions and in July 2001 LocusDialog was formed around the enterprise solution product, known as Liaison. InfoSpace is focused exclusively on the service provider environment, while LocusDialog will primarily be a CPE supplier for enterprises.

LocusDialog was purchased by some of the original Locus Dialogue investors and the only link that now exists between InfoSpace and LocusDialog is that InfoSpace is the supplier of the base ASR technology to LocusDialog (InfoSpace retained ownership of this). LocusDialog views this as a viable solution, since they still possess the technology, but do not face any further base technology development costs. Plus, it gives LocusDialog the flexibility to pursue a best-of-breed strategies in the future as they advance in the enterprise space.

Locus Dialogue was originally founded in 1996, as a spin-off of the Computer Research Institute of Montreal (CRIM), whose founders developed CRIM's speech recognition technology. In total, Locus Dialogue's core technologies and applications have been in development for more than 14 years and have earned numerous industry awards, including honors from Computer Telephony Expo 2000, Communications

Solutions Expo 2000, and the 2001 Multimedia International Market Grand Prize for Business Solutions.

With its flagship Liaison product, Liaison, LocusDialog is a leading supplier of speech-enabled call routing systems with an installed base of approximately 500 systems, which according to LocusDialog represents about 24 percent of the total worldwide installed base. Customers include: Zurich Insurance, Daimler-Chrysler, Deloitte&Touche, HarrisBank, McKesson, ReMax and Kemper Insurance. LocusDialog has developed a strong distribution channel in North America and is now in the process of further expanding its distribution outside of North America. The LocusDialog distribution channel is comprised of leading telecommunications equipment suppliers. As LocusDialog expands further outside of North America, the capability it has of utilizing other ASR technology for other languages could prove to be attractive.

LocusDialog's Liaison system is a speech-enabled call routing solution that provides a variety of telephony applications for enterprises, including call answering, call center routing, corporate and personal voice-activated dialing and the provision of self-service information. Callers into a corporate directory that uses Liaison may simply state their request by voice and the system routes them accordingly. The Liaison system delivers speech functionality through integration with PBX telephony systems. The LocusDialog Liaison system version 6.0 is offered in three editions:

1. The Liaison Lite provides 24/7 speech-enabled call routing for small organizations and medium-sized enterprises (still in beta phase; January 2002 launch);
2. Liaison Enterprise is designed for small and medium size organizations; and
3. Liaison Enterprise XL is designed specifically for large firms with distributed locations.

The Enterprise and Enterprise XL editions incorporate enhanced features, speech recognition capabilities, and offer new, optional application modules. New standard features include: Transfer Entry which provides for up to 8 separate call destinations; Call

Redirect which provides call redirection to 8 phone numbers; and Dynamic Call Routing which uses DNIS and CLID to dynamically routes calls to different menus.

### Key Developments

- In October 2001, LocusDialog announced a partnership with EADS TELECOM, a supplier of secure communications solutions for defense, public safety and civil applications. EADS TELECOM is partly owned by Nortel. LocusDialog will speech-enable EADS TELECOM's Intelligent Directory System, which is a PC-based software console attendant with a Web directory interface. It allows users of the system to automatically route their calls to the requested person by simply speaking the name rather than going through an operator.
- Datapulse, a supplier to British Telecommunications is including LocusDialog Liaison in its CRM product suite. Datapulse works with telephony system manufacturers in the UK. With the addition of the Liaison system to its suite of product offerings, Datapulse will be able to bring speech recognition to a long list of clients in the UK and Europe. Distribution is further extended beyond the borders of the U.K. through Datapulse's European partners Tele Danmark (Denmark), InTime Netbuilding (Netherlands), Compta (Portugal), TCC (Germany), and Netser (Turkey).

#### **3.3.1.9 Lucent Speech Solutions**

Lucent Bell Labs has experience in the speech technology area; it has been granted some 140 patents in speech processing over the years. This technology base has been integrated into products of Lucent Speech Solutions, which was a new division created on January 28, 1999. Its product line consists of three core products: the Lucent Speech Server, Compact PCI Speech Processing boards, and the Lucent Articulator TTS software. First introduced in April, 1999, Lucent's speech technology network server, called Lucent

Speech Server, provides an integrated telephony/speech solution that supports a range of speech-enabled solutions, like voice activated dialing, auto attendant, and personal assistant applications. The Lucent Speech Server is a complete network-grade platform that enables developers to create and host speech applications on one integrated system that includes hardware, software, telephony, speech processing engines and operation support systems. Developers can program to the Speech Server's Java interface, which includes a library of application level Speech Actions. The Speech Server is built from Lucent Compact PCI Speech Processing Boards that feature a full suite of speech resources (ASR, TTS, DTMF, Play, etc.). The server supports up to 192 channels and is able to support both ASR and TTS functionality simultaneously.

Lucent Speech Server product is ahead of the marketplace as a turnkey solution since it includes the telephony application, management tools and speech technology functionality all in one package. This includes valuable middleware, administrative resource management tools and VoiceXML interpreters. Another key advantage is that it can be scaled up to support more than one million users per system, since clusters of Lucent Speech Servers can be dispersed globally, yet can be managed as one unit. In early 2002 Lucent will introduce the version 300 of Lucent Speech Server, which is a smaller version (four T-1 interfaces) of the main server, designed for smaller operations..

Also for the lower end of the market, Lucent offers its Text-to-Speech engine as a Software Development Kit for Windows and Unix-based platforms, called Articulator. Lucent Articulator was launched April 2, 2001 and is geared toward users who need to support traffic of 48 channels or less. Articulator contains the same Bell Labs technology found on the speech processing boards used in its Speech Server. The SDK includes speech models optimized for desktop, email and telephony applications with one of its chief applications being speech-enabled access to email from a remote location, such as mobile phones. It includes support for the following languages: American English, German, Mandarin Chinese, Spanish, French, French Canadian, Italian. More than 25 customers are currently in trial with the new speech software solution. Beta customers in trial include Nuance Communications, Philips Speech Processing, Motorola and Sound

Advantage. Lucent is also deploying the Articulator software as an enabling technology in the Lucent MiLife Media Platform and the e-Services product portfolio.

#### Key Developments

- Lucent Technologies has a new customer, MetroCommute.com, who will use automatic speech recognition and TTS technology from Bell Labs to provide traffic information over the telephone. As part of a new business agreement, MetroCommute.com has acquired the worldwide rights to license Lucent's PhoneBrowser technology, which is Lucent's voice portal software. MetroCommute.com, based in NYC and founded in 1994, currently provides real-time traffic information through its Web site, [www.metrocommute.com](http://www.metrocommute.com), via e-mail alerts to subscribers' desktops or mobile phones, and as data and video feeds to television and radio stations.
- Sound Advantage and Lucent Technologies formed a technology alliance to incorporate speech technology into the Sound Advantage line of unified messaging systems. SoundAdvantage is also a major e-mail reading provider and will use the Labs' high-quality TTS to improve its product offerings. Sound Advantage products serve the wireless and telephony markets. Lucent's technology will be the next version of SANDi (Sound Advantage Natural Dialog Interface), which has won some 20 industry awards. SANDI is used by enterprises to support voice-activation of all office communications including: answering calls, transferring calls, locating employees and handling unified messaging.

#### **3.3.1.10 Microsoft Speech.NET**

Microsoft Speech.NET was formed in September 2001 and combined both the speech research and product groups. In the past, Microsoft focused on the desktop: evolving speech technology into speech-enabled applications for the Windows platform. This

resulted in development of Speech Application Programming Interface (SAPI), which is now version 5.1. Now, they are designing a platform that will support other devices well, such as wireless PDAs, multimedia handheld devices like MiPad and the standard mobile telephone. The Speech.NET vision is to make it possible for people to access information by speech anywhere, anytime, on any device. Speech.NET just finished and recently launched speech-enabled components in the new Office XP and Windows XP. Both contain SAPI, which contains a basic speech-to-text module and development tools for doing more.

The Microsoft.NET Speech SDK comes with three developer tools that provide interfaces through which comprehensive speech-enabled Web applications can be created. Web developers can configure their Web applications for multimodal applications and voice-only access. Multimodal access will enable users to interact with an application in a variety of ways: they will be able to input data using speech, a keyboard, keypad, mouse and/or stylus, and produce data as synthesized speech, audio, plain text, motion video, and/or graphics.

The major discussion about Microsoft at SpeechTEK 2001 was its move to propose and move forward on a new standard, called Speech Application Language Tags (SALT) that would be an alternative to VoiceXML. Microsoft's approach to speech enabling the Web is designed to eliminate the need for converting HTML to VoiceXML. In Microsoft's view, SALT is better since the same programming code can be used for Web and speech-enabled versions, so they can be built at the same time. The new markup language extends HTML with a small number of elements and objects that add speech recognition input, audio and text-to-speech playback, along with telephone call control to the Web application.

Current Microsoft speech tags included in its new launch of Windows are built around what it hopes will be the new emerging standard, SALT. A newly formed industry organization, called the SALT Forum, will produce version 1.0 specification of SALT and submit it to the standards bodies. The SALT Forum is committed to developing a royalty-free, platform-independent standard that will make possible multimodal and telephony-

enabled access to information, applications, and Web services from PCs, telephones, tablet PCs, and wireless personal digital assistants (PDAs). The new standard will extend existing mark-up languages such as HTML, XHTML, and XML.

### **3.3.1.11 Motorola**

Motorola's goal in speech technology is consumer focused. Its goal is to provide leading-edge solutions that allow consumers to access Internet content and services through the power of voice virtually anywhere in the world. To that end, the company ensures that it works with leading technology and component vendors, such as SpeechWorks and others. The MVP Systems & Software unit in Motorola has developed a set of speech technology platforms, called Mya. The Mya Voice Platforms (MVP) are enabling voice-enabled access to the Internet through powerful, integrated systems that simplify and streamline development, deployment and usage of the voice Internet. The heart of MVP is Motorola's field-proven Voice Browser. Announced in June 2000, the VoiceXML-compliant browser translates ASR outputs from human speech, captures requested information from voice-enabled websites, and delivers it to TTS software that reads the relevant data to the end user.

On August 6, 2001, SpeechWorks and Motorola announced that Motorola added SpeechWorks' Speechify and ETI-Eloquence text-to-speech technology to its voice portfolio. Under the terms of the agreement, Motorola will license SpeechWorks' solutions for use in its commercially available Mya Vx500 Voice Platform. This agreement, combined with Motorola's relationships with other vendors, will give wireless carriers and enterprises a broadest selection of TTS solutions in the marketplace. SpeechWorks' Speechify software converts text input from any source into high-quality speech that is easy for callers to listen to and understand. The Company's ETI-Eloquence TTS engine has multi-language/dialect capabilities that make it a functional solution for companies operating around the world. In addition, ETI-Eloquence has a very small footprint, making it ideal for use in embedded devices. Both products provide high-quality speech



output that enhance the caller's experience by providing dynamic, detailed and personalized information in an automated, near-human voice.

### Key Developments

- Verascape, Inc., a manufacturer of high-capacity VoiceXML speech platforms, and Motorola formed a partnership in which Motorola will resell Verascape's UNIX-based VeraServ speech products under a Motorola private label. The announcement comes a few months after a licensing agreement whereby Verascape integrated Motorola's Mya VoiceXML browser into its VeraServ platform. This collaboration will enable Motorola to expand its Mya Voice Platform product portfolio to provide the industry's most comprehensive range of NT and UNIX-based speech infrastructure products to domestic and international carriers for writing or hosting voice-based telephony applications in VoiceXML. This collaboration also enables Verascape to leverage the large developer base that Motorola has cultivated for its industry-leading Mya VoiceXML browser, as well as reach carrier markets through the strength of Motorola's carrier distribution network.

#### **3.3.1.12 Nuance Communications**

Nuance's origins are with SRI International, which was one of the leading voice technology research entities throughout the 1980s and 1990s. Nuance was formed as an independent company by SRI in 1994. Nuance investors include Alloy Ventures, Asset Management, Mayfield Fund, Morgenthaler Ventures, and USVP. Corporate investors include Cisco Systems, Intel, Motorola, SAIC, Siebel Systems, SRI International, Sun Microsystems and Visa International. Nuance offers its products through industry partners, platform providers, and value-added resellers around the world. Major partners

are Nortel Networks and Tellme. Major customers include Sears, American Airlines and Charles Schwab. Other important customers of Nuance are:

- Fidelity
- Lloyds TSB
- Sprint PCS
- United Parcel Service
- Verizon

Nuance (Menlo Park, California) was one of the first companies to market voice browser products and is certainly one of the most active players in the speech recognition market. Nuance 7.0, with its V-Builder and Voyager platform products, is the main offering from Nuance Communications. Its software recognizes twenty-six languages and even recognizes when a bilingual caller switches languages or uses a different pronunciation of a name or dialect. Nuance supports these 26 languages:

- Australian-New Zealand English;
- Brazilian Portuguese;
- Canadian French;
- Chinese Cantonese;
- Chinese Mandarin (China);
- Chinese Mandarin (Taiwan);
- Czech;
- Danish;
- Dutch;
- European French;
- European Spanish;
- German (Austria-Germany and Switzerland);
- Greek;
- Hebrew;
- Italian;
- Japanese;

- Korean;
- Norwegian;
- Singapore English;
- South African English;
- Spanish;
- Swedish;
- Turkish;
- UK English; and
- US-Canadian English.

The latest versions of its products were specifically updated to work in wireless environments. Speech recognition rates have been measured at over 96 percent accuracy on their customers systems for Charles Schwab & Company, Sears and Odeon. Nuance has created a Developer Network that already serves more than 10,000 developers. The Developer Network helps developers with software development kits and the tools they need to develop voice portal applications. They also offer the Voice Web ASP program that assists ASPs in hosting and delivering voice-enabled applications of all kinds. Its Voice Web Content program offers consulting and syndicated content for dissemination by multiple portals.

The upgraded version of Nuance's speaker verification product was announced in the second half of 2001. Nuance Verifier 3.0 can eliminate security breaches for airlines, enterprises and government agencies through voice authentication. Compared to other biometrics, including fingerprinting, handprinting, iris, retina or facial scanning, voice authentication was proven as a superior security solution, showing a 99.9 percent accuracy rate in a recent research study commissioned by Communications Electronics Security Group (CESG) in the UK. Nuance Verifier performs voice authentication from any network access point with a microphone or ordinary telephone. No special equipment is necessary and the technology can handle a range of applications from facilities access to the protection of sensitive data. Verifier 3.0 now offers enhanced speaker identification which enables voice authentication systems to better identify individual

members of groups, such as families or small offices, that may share a single account; greater accuracy – improved performance in noisy environments, such as cars or airports; and superior name recognition – a tighter integration with Nuance speech recognition software improves the identification of all individuals when they speak their names, even if multiple users have the same name.

Nuance Verifier is proven technology for enhanced security, as demonstrated by global deployments, including United Kingdom government's Intensive Supervision and Surveillance Programme for fighting crime; and major financial institutions for access to personal accounts and information. A range of companies have integrated Nuance Verifier into their solutions, including Aspect Communications, Edify, InterVoice-Brite, Mitel, Nortel Networks, Premier Geografix and Vocent.

### Key Developments

- In December 2001, AT&T Labs announced it will integrate its text-to-speech software, called Natural Voices, with Nuance's Vocalizer Integration Framework, TTS engine integration platform.
- Nuance's Vocalizer Integration Framework (VIF) for the text-to-speech market was launched. VIF consists of software that enables the integration of the best-of-breed TTS engines from Nuance and other vendors with a patented Nuance distributed architecture. The framework integrates TTS capabilities with the Nuance 7.0 speech recognition engine and the Nuance Verifier 3.0 voice authentication software. Nuance distributed architecture enables a speech system to load balance between servers to best use all available hardware resources.
- Oracle Corporation announced support and deployment of voice-enabled services across its platforms. This includes Oracle9i Application Server Wireless, which will use ASR and TTS technologies from Nuance and SpeechWorks. It opened a voice technology development center in Chicago.

- Nuance announced an enhanced product line that supports voice over Internet protocol (VoIP) including support for the standards: Session Initiation Protocol (SIP) and H.323. Nuance added Net2Phone, a provider of VoIP services, to its partner network, which already includes VoIP developer's iBasis, Indicast, SandCherry Networks and Voxeo.

\* Rhetorical Systems, based in Edinburgh, will integrate its text-to-speech software, called rVoice, with Nuance's Vocalizer Integration Framework, TTS engine integration platform.

### **3.3.1.13 Philips Speech Processing**

Along with Lernout & Hauspie, the speech recognition and natural language understanding product portfolio of Philips Speech Processing is certainly one of the most comprehensive and technically advanced in the market. Philips has developed its core technology over the past 30 years. Philips' speech recognition unit operates as a business unit of Philips Consumer Electronics. Recently this Philips unit, with headquarters in Germany and US operations out of Dallas Texas, shifted its attention from consumer dictation software for the PC to telephony applications. Philips also expanded its speech portfolio last year by acquiring Voice Control Systems (Dallas, Texas) for \$60 million.

Philips offers three major products: SpeechPearl 2000, SpeechMania, and SpeechWave. SpeechPearl 2000 is its core speech recognizer product, which includes natural language software that understands continuous speech. SpeechMania is a speech recognition developer environment and SpeechWave is a chip-based solution for handset vendors.

On July 16, 2001, Philips announced that its speech recognition technology is now available in a further ten languages. Philips Speech Processing has one of the largest portfolio of languages and accents of any vendor in the speech recognition market. The latest language additions are:

- Russian;

- Czech;
- Hungarian;
- Slovakian;
- Slovenian;
- Catalan;
- Welsh;
- Cantonese;
- Flemish; and
- Luxembourg German.

Much of Philips' focus is outside the US, with deployments covering a wide range of applications. SFR, a French wireless company that is part of Cegetel, launched a voice portal called VOX using Philips technology. Philips Speech Processing has worked with Omnitel Pronto Italia, the second largest mobile operator in Europe, to launch Omnitel 2000. It is the largest implementation of a voice portal and a natural language application in an intelligent network of any in the world. Using Omnitel 2000, mobile phone are able to access more than 300 different data bases, including restaurant guides, yellow pages, white pages. Early reports on usage are very encouraging.

During the year 2001, Philips extended sales in the voice dialing arena, which it considers one of the markets with the greatest potential. Here in the US, Philips has licensed its technology to Preferred Voice, a Dallas-based applications service provider, which hosts and manages voice dialing services for wireless firms. Philips has authorized Preferred Voice's customers to use Philips' patented voice dialing technology. The firms are:

- Golden State Cellular;
- Rural Cellular Corp;
- Midwest Wireless;
- Illinois Valley Cellular;
- Cellular One of Illinois;
- XITCommunications;

- Pioneer Enid Cellular;
- Cell XL;
- Highland Cellular;
- Cellular Mobile Systems; and
- Thumb Cellular.

### Key Developments

- Philips Speech Processing is working with Respons AB on creating Autosvar, Sweden's new speech-enabled directory assistance service. Respons AB, a subsidiary of Industri Kapital, a UK-based private equity investment firm, has decided to launch a more affordable alternative to traditional operator-assisted service. Autosvar is Sweden's nationwide service for residential and business listings that can be reached by dialing 118 888 within the country. Autosvar deploys Philips' SpeechMania 7.0 speech recognition platform and the directory assistance module SpeechFinder.
- Telefónica de Argentina announced the launch of first two speech-enabled DA services driven by Philips' speech technology. The two services are fairly simple; the first service provides callers with the telephone area code for a city or country. And the second service enables users to receive a new telephone number – after stating the old number – for a person or business whose number has changed. These are two initial services will be extended later by the launch of Telefónica de Argentina's full speech-enabled directory assistance service portal. This portal will be accessible by a single telephone number and will allow users access to speech-enabled directory listings and services.
- Rhetorical Systems will make its text-to-speech software rVoice available to Philips. rVoice, Rhetorical Systems' core product, transforms written words or texts into natural, pleasant sounding speech. This added-value product

complements Philips' speech recognition applications by adding a very high voice quality to TTS.

### **3.3.1.14 Phonetic Systems**

Founded in 1994, Phonetic Systems of Bedford, Massachusetts offers voice-enabled directory systems for large corporate and enterprise customers. In July 2001, Phonetic Systems appointed a new CEO -- John D. Howard. Mr. Howard was the former Nortel Networks Executive who led its China wireless division. Phonetic Systems' solutions are based on proprietary, patented platforms, including its Voice Search Engine (VSE). Phonetic Systems dominates the US market for voice-enabled corporate directories, with over 100 customers among Fortune 1,000 companies using its VSE technology for such applications as speech-enabled auto attendants, call routing and directory information centers. Customers of PhoneticOperator include:

- Atlantic Records
- Bear Stearns
- Credit Suisse
- Fannie Mae
- McGraw Hill
- Microsoft Corporation
- Motorola
- Price, Waterhouse, Coopers
- Stanford University

Phonetic Systems upgraded its flagship product PhoneticOperator to version 4.0. PhoneticOperator is described as a voice-driven auto attendant, call routing and information retrieval solution. The new version provides a new scalability that offers voice access to databases from a few thousand records to over one million. The system offers real-time updates to large corporate directories. Other key enhancements are:



- Multi-language – supports additional European languages, allow easy installation on top of existing PhoneticOperation installations;
- Digital Telephony Integration – digital telephony integration to PBX and large corporate switches, monitoring for busy signaling, dynamic port allocation, and use of ANI/DNIS information;
- SNMP – enables management through a corporate network management station, providing complete network and system software status.

Phonetic Systems reported record sales for second quarter, 2001 with significant increase in total number of new orders, and expansion of orders from existing customers.

The key partnerships it has developed are:

- Persay Ltd, a Converse Technology spin-off, to integrate its voice verification systems into Phonetic System products
- VoiceGenie Technologies, a provider of VoiceXML Gateway solutions. Partner on voice Web solutions.
- Synamics, provider of integrated voice and data communications solutions. Integrate Phonetic Systems' voice directory with Synamics's Telephony Application and Media Server
- Business Layers, a provider of e-commerce solutions, integrate with Business Layers' eProvision solution

### Key Developments

- Phonetic Systems launched a free telephone directory assistance service offering access to 700,000 phone numbers for Eastern Massachusetts businesses through a system that relies entirely on speech recognition. Privately held Phonetic said it hopes the toll-free number, 866-8-FREE-DA (866-837-3332), will help demonstrate to big phone companies the potential for automated directory assistance. It also expects to get valuable real-world experience with the system to work out glitches. Working with LSSI, a New Jersey-based information company that maintains a

database of 175 million US phone numbers, Phonetic has assembled listings of 700,000 business numbers in the four Eastern Massachusetts area codes for its "Free D.A." number.

- Phonetic Systems and InfoSpace struck a joint agreement to provide non-operator assisted, speech-enabled directory service applications for wireless carriers and voice service providers. The directory service applications will also extend the reach of InfoSpace's base of online, offline, local, national and international merchants. The applications make use of Phonetic Systems' ability to voice-enable large and dynamic directories, and InfoSpace's wireless and Internet software and application services. Phonetic Systems' patented Voice Search Engine (VSE) technology, which scales to voice-enable millions of records allowing users to rapidly access information from virtually any size directory, is planned to be combined with the InfoSpace Speech Platform and speech-enabled directory service applications. The directory service applications allow users to request residential or business listings by simply speaking their request to a speech technology system.
- Phonetic Systems established a partnership with VoiceGenie Technologies, Inc., the leading provider of VoiceXML Gateway solutions, to offer customers Voice Web solutions that are 100% compliant with the industry standard. Customers will now have powerful and easy-to-use voice access to unlimited Web information. This partnership enables Phonetic Systems to integrate its leading-edge voice automated directory products, like PhoneticOperator, to offer customers access to VoiceXML-based information, conduct on-line transactions, and manage personal communications, including e-mail and voice-activated dialing, simply by using their voice over any phone, anytime, anywhere. In teaming with VoiceGenie, Phonetic Systems will deliver VoiceXML compliant solutions that enable voice-driven access to information from a directory of millions of records accurately and instantly.

### **3.3.1.15 ScanSoft**

ScanSoft produces digital imaging software and competes with the likes of Adobe Systems. About three-quarters of its revenue (\$64 million in 2001) comes from three products: an optical character recognition device, TextBridge; a scanning and editing images product, OminiPage Pro; and scanned document manager, PaperPort. Located in Peabody, MA, ScanSoft is 25 percent owned by Xerox. In late 2001, ScanSoft bought some of the key assets of Lernout & Hauspie's speech products for \$40 million during the bankruptcy auction.

The acquisition of L&H's Speech and Language Technology assets was completed December 12, 2001. They hired about 225 key employees in R&D, sales and marketing. ScanSoft has kept the name Dragon NaturallySpeaking for its dictation software, which it signed license agreements with AT&T, Chevron and Kodak. The market for Dragon is seen for corporate, healthcare and legal markets. It signed new agreements for its embedded speech technologies, also acquired from L&H, with Nortel, Teleco and Vodavi Communications.

Going forward, ScanSoft looks to the telematics market, where they are marketing to customers in North America, Europe, and Asia. Total revenues are expected to grow to \$105 for the year 2002. Opportunities in speech and language solutions are expected to fuel much of this growth.

### **3.3.1.16 SpeechWorks International**

SpeechWorks (Boston, Massachusetts) is the other speech technology powerhouse in the US, besides Nuance. SpeechWorks is the market leader in the US based on revenue. Founded in 1994, SpeechWorks has over 400 employees worldwide and works with over 100 partners to implement speech solutions in leading corporations, telecommunications firms, automotive and device manufacturers, and government agencies. SpeechWorks has recently launched a major initiative to foster its speech technology on the VoiceXML standard. The products are called OpenSpeech.

OpenSpeech embraces a whole set of products including the OpenSpeech Recognizer, which is SpeechWorks' core ASR engine, OpenSpeech DialogModules that permits easy implementation of standard applications, and OpenSpeech Browser Platform Integration Kit, a developers' kit based on SpeechWorks' OpenVXI.

In essence, the OpenSpeech product line advances the goal of expanding use of speech technology by widening the available set of platforms that fit SpeechWorks software. OpenSpeech Server software allows customers to operate their speech applications in a client-server, standards-based environment. OpenSpeech Server works in conjunction with SpeechWorks' OpenSpeech Recognizer software, an open, standards-based speech recognition engine that is optimized for VoiceXML 2.0. OpenSpeech Server creates a bridge between the recognition engine and the telephony client, for example, for those using an IP network. This architecture allows a call that reaches the telephony client to be assigned to one of a number of speech recognition servers that communicate across IP networks using standard protocols, giving the carrier or service provider greater flexibility in the choices they make in speech technology.

During 2001, SpeechWorks has expanded or upgraded its product offerings to include:

- SpeechWorks 6.5 SE -- an enhanced version of SpeechWorks speech recognition engine and associated tools, offering high accuracy in wireless environments and the ability to improve accuracy by automatically adapting to caller usage in live systems;
- SpeechWorks Speechify -- an advanced text-to-speech (TTS) product that uses advanced AT&T Labs research to market delivering pleasant, understandable and natural-sounding synthesized speech;
- SpeechWorks ETI-Eloquence -- a powerful and efficient TTS system with multi-language capabilities (13) and low memory requirements ideal for embedded devices such as in-car navigation systems, next generation mobile phones and hand-held devices;

- Speech2Go – is a small footprint, embedded speech recognition engine designed specifically for use in automotive and mobile devices. It supports Distributed Speech Recognition.
- SpeechWorks SpeechSecure – speaker verification, a biometric technology that verifies a caller's identity based on the characteristics of his or her unique vocal patterns; and

\* SpeechSite -- a packaged application that enables corporations to greet callers, route their calls and respond to requests 24 hours a day.

SpeechWorks has done much to advance creative speech-enabled solutions for call centers. A number of major firms have used SpeechWorks to upgrade their toll-free response centers with its technology. For example, Amtrak launched a new voice-enabled system for customers to check train status on its toll-free reservation line. The nationwide system uses SpeechWorks technology to serve the 2.7 million calls that come into 800-USA-RAIL each year that request train status information. Using natural language speech recognition, the system understands the caller's words and provides answers on command. Customers do not need to know number of the train, but by using origin/destination cities and arrival times, users can obtain the status of any train in the US using spoken commands. Amtrak has stated that the system has paid for itself in just a few months and the customer response has been very favorable. Some further examples:

- SpeechWorks helped Thrifty Car Rental and its call center. Thrifty callers can shop for the best available rental rates through the main Thrifty reservation number, 1-800-THRIFTY.
- First Union Wachovia has deployed the speech-enabled largest banking app in the US with SpeechWorks. The new service allows callers to conduct transactions and get customized banking information.
- E\*Trade has enhanced its main number at 800-STOCKS1, where customers can use a speech-enabled telephone trading service.

- AOL uses speech technology for the dial-in version of its Web information, called AOLByPhone. With partner Quack.com, SpeechWorks developed a voice portal for American Online's 24 million customers.
- For United Airlines, SpeechWorks has succeeded in creating a model call center deployment. At UA's 1-800-824-6200 number, callers can check the status of any of United's 2,400 daily flights by using spoken commands. UA reports that over 80,000 people now use the system each day to obtain automated flight information by voice.

Some of SpeechWorks' other current customers are: Continental Airlines, Flight Information System; Crédit Lyonnais FedEx; Fidelity Investments; HeyAnita; Hyundai Securities; Manulife Financial; McKesson; Office Depot; and TD Waterhouse Australia.

SpeechWorks has a leadership role in several standards bodies including; ETSI (European Telephony Standards Institute) Aurora committee, which is developing an industry standard for distributed speech recognition; SALT (Speech Application Language Tags) Forum, which was formed to accelerate standards for multimodal applications; and, the VoiceXML forum of which SpeechWorks has been an active participant on four committees. SpeechWorks engineers have helped create and edit the SSML and VoiceXML 2.0 specifications.

### Key Developments

- SpeechWorks formed a new business unit to meet the growing demand for speech applications in the automotive, mobile device and set-top box industries. The newly created Automotive and Mobile Device Customer Focus Unit provides dedicated product and business development as well as marketing and sales resources to meet the speech recognition and text-to-speech demands of customers in these industries. SpeechWorks partners in the device market include America Online (NYSE: AOL), Compaq (NYSE: CPQ), Kirusa, LOBBY7, Mitsubishi

MERL (Mitsubishi Electric Research Laboratories), Motorola (NYSE: MOT), OnStar, Texas Instruments (NYSE: TXN), Visteon and Wingcast.

- SpeechWorks established its Japanese subsidiary, SpeechWorks Japan KK with offices in Tokyo, Japan. SpeechWorks Japan will address the growing market demand for speech solutions in Japan, because there is a significant market for speech-enabled services in the Asia Pacific region and especially in Japan.
- Nortel Networks announced a partnership with SpeechWorks late in 2001. Nortel will combine its OSCAR IVR system with SpeechWorks' OpenSpeech Recognizer to offer enterprise and carrier customers a joint IVR/speech recognition solution that will be launched during first quarter 2002. Nortel is targeting applications such as order fulfillment, flight information, and banking transactions for the financial, travel, insurance, telecommunications and healthcare industries.
- National Weather Service, an agency of the Commerce Department's National Oceanic and Atmospheric Administration (NOAA) selected SpeechWorks' Speechify TTS software for the next-generation voice of the nation's automated radio weather warning system which is broadcast on 583 stations across the nation, reaching 85% of the US population. The NOAA awarded systems integrator and SpeechWorks partner, Siemens Information and Communication Network of Boca Raton, Florida the contract. The weather service will begin implementing the new system in early 2002 with future enhancements to include Spanish language capabilities. The old and new voices can be heard at <http://www.nws.noaa.gov/nwr/newvoice.htm>
- Unisys is integrating the SpeechWorks OpenSpeech Recognizer into its extensive Unisys [e-@ction](#) Multimedia Value-Added Services. The OpenSpeech Recognizer is optimized for VoiceXML, which complements the Unisys strategy to provide open speech-enabled solutions to customers worldwide based on Unisys scalable platforms, integration skills and best-of-breed components. Unisys [e-@ction](#) Multimedia Value-Added Services (MVAS) is a solution set that encompasses voice/fax messaging, universal messaging, calling services, Internet call waiting,

and other enhanced messaging services. This announcement is significant since Unisys supports more than 100 million voice/fax mailboxes for over 100 MVAS clients in 40 countries.

### **3.3.1.17 SRC**

SRC is one of the top UK-based provider of high-performance speech technology solutions for telephony, desktop, and enterprise solutions. SRC has a long track record of working with major UK companies, and has delivered speech solutions to over 20 FTSE 100 organizations [note: FTSE is a leading global stock indices in the UK and Europe]. SRC is privately owned, and is backed by leading financial institutions including 3i Group and Cazenove & Co. Some of SRC's key customers are: Adams & Remers; Arthur Andersen; British American Tobacco; British Telecom; Deloitte & Touche; HSBC; Morgan Cole; Morgan Stanley Dean Witter and Vodafone

SRC provides both desktop and telecom based speech solutions. SRC Enterprise delivers desktop and enterprise solutions to organizations that require speech dictation, custom vocabulary building, bespoke speech application development, digital speech workflows or a mixture of these. SRC Telecom provides hosted telephony speech recognition services and a complete range of speech recognition applications implementation and support services. SRC also provide consulting services advising on all aspects of utilizing speech technologies for business advantage.

In January, SRC formed SRC Telecom, which will be a division of SRC focused on developing and delivering telephony-based speech recognition solutions. Their services are targeted at companies that need to improve their call centers, at voice portal offerings and voice commerce. Dr. Neil Bowers, previously from Canon Research Europe, will be leading the efforts at developing telephony applications. Mid-year SRC launched its VoiceXML applications hosting service. SRC is leading the introduction in Europe of VXML platforms that allow third party developers to create telephony applications using its hosting environment. Based on SRC's speech platforms, the hosting environment can



aid companies develop their own custom speech applications based on the VXML markup language.

The other major unit in SRC related to telephony is SRC Enterprise. It provides speech solutions, such as call routing, to major UK businesses. One of its customers is Baker Tilly, a top 10 accounting firm. Two major legal firms, Irwin Mitchell and Morgan Cole, are rolling out an SRC speech system in their firms. Other firms are:

- Redfern & Stiggant
- Le Brasseur J Tickle
- Kennedys
- Gartmore Investment Management

#### Key Developments

- VoiceGenie Technologies, a key provider of VoiceXML Gateway services, is now partnering with SRC. SRC will use VoiceGenie's VoiceXML Gateway to deliver telephony-based speech-recognition services. SRC Telecom will deliver its offering in an ASP environment. The VoiceXML Gateway provides SRC Telecom with an E1 base running on the UnixWare operating system. This announcement is significant because it marks VoiceGenie's entry into the European marketplace. The platform can be used to support clients' customer-service initiatives or to enable clients to offer speech-based information and transaction service.
- Cap Gemini Ernst & Young, management consulting and IT firm, is partnering with SRC on telephony-based speech recognition. This is a significant agreement since Cap Gemini is the second largest systems integrator in the telecommunications world. SRC will combine its speech solutions with Cap Gemini's CRM business to offer telephony-based speech-enabled applications for the call center market. SRC platform will integrate the Nortel MPS platform with best-of-breed speech technologies, including those from SpeechWorks and Nuance.

#### **3.3.1.18 Telisma**

Telisma was created in August 2000 to provide solutions based on speech recognition technology initially developed at France Telecom R&D. France Telecom became a minority shareholder, along with other partners Innovacom, AGF Private Equity and 3i. The company now has a staff of 63. Telisma technologies have already been implemented in the following situations:

- France Telecom's telephone information service – France Telecom has created a Directory Assistance Call Completion (DACC) service using Telisma's speech recognition engine.
- Orange messaging service – Orange's voice mail system, Mobicarte, lets users consult their voice mail by dialing 777 and using voice commands to listen to messages.
- Voice portals – Telisma has assisted France Telecom and Orange to create voice portals. The Orange 711 portal now offered to its 20 million customers and lets them obtain information and conduct transactions, such as book a taxi, listen to their horoscope, find movies, and deliver flowers.

Telisma focuses on making speech technologies an integral part of the telecommunications and I.T. infrastructures, in order to enable truly ubiquitous, pervasive, and intuitive access to information for businesses and consumers alike, especially in a mobile environment.

Telisma partners include a number of key international players. They are portrayed in the following chart that shows the partner and the arena in which they play.

Table 3.x

Telisma Partners

	Chips	Systems	Telephony boards	TTS	Telephony Platform	Integrator	Access Providers
Alcatel					X	X	
Atos					X	X	X

Bull						X	
Cap Gemini					X	X	
CVF						X	X
Elan				X			
HP		X			X		
IBM		X			X	X	
Idylic					X		
Intel			X				
NetCentrex					X		
NMS		X	X		X		
Prosodie							X
T.I.	X						
Unisys					X	X	
WellX					X		

Telisma has instituted a range of products since its recent launch. They include speech recognition software, development tools. Their preferred TTS provider is Elan. Their products are:

- Telisma Philsoft, v.3 – Telco-grade, distributed automatic speech recognition engine, supporting multiple languages, including US English, UK English, French, Spanish, Italian, German, and more. Philsoft has several options, including a Dynamic Grammar Builder, a module enabling development of speech verification applications, and more;
- Telisma Philmodeling Studio -- grammar development & modeling tool, graphics-based;
- Telisma VoiceXML Interpreter – optimized, telco-grade interpreter;

- Telisma VoiceXML Evaluation Kit – a VoiceXML platform, including the Telisma VoiceXML Interpreter, Telisma Philsoft for speech recognition and TTS from Elan Informatique.

### Key Developments

- Telisma and Eureka Soft, an integrator specializing in telecommunications platforms for the creation and management of Added Value Services, announced the integration of Telisma Philsoft® v3 and the porting of Telisma's VoiceXML interpreter into the Eureka Soft Orateur Added Value Service creation environment. This announcement highlights the desire of the two companies to combine their expertise and their extensive technological experience to provide European fixed and mobile operators and speech-based service access providers with a complete Added Value interactive voice service offer
- Telisma has entered a partnership agreement with Thomson multimedia (TMM) by which Thomson multimedia becomes a shareholder of Telisma and Telisma gets a license on TMM's patents in the speech recognition field. The TMM R&D team for speech recognition is transferred to Telisma and both companies will cooperate on the development of speech-enabled applications for consumer products.

#### **3.3.1.19 Unisys**

Unisys calls its speech technology efforts the Unisys Speech Solutions Consulting Practice. Actually speech technology solutions are part of a much larger telephony and multimedia applications effort, called Unisys [e-@ction](#) Communications Solutions. This is a broad portfolio of standards-based, end-to-end multimedia applications specifically designed to meet carrier-class requirements for scalability, availability, interoperability, security, and customizability. Unisys has developed solutions for voice activated dialing, network White Pages, enterprise VAD, and voice portals. Unisys has won over 100

communications service provider customers (including 9 of the 10 world's largest) in 40 countries. The key segments of this effort are divided into a number of areas, including speech recognition:

- Multimedia Value Added Services including Voice/Fax Messaging, Unified Messaging, Prepaid Card, Prepaid Mobile, and Multimedia;
- e-Telco (OSS/BSS) Solutions include Remittance Processing and Call Centers/CIS;
- Broadband solutions include Video on Demand;
- Mobile Business Solutions; and
- Natural Language Speech Recognition solutions.

#### Key Developments

- Unisys announced late in 2001 that it was integrating the SpeechWorks OpenSpeech Recognizer into its extensive Unisys [e-@ction](#) Multimedia Value-Added Services. The OpenSpeech Recognizer is optimized for VoiceXML, which complements the Unisys strategy to provide open speech-enabled solutions to customers worldwide based on Unisys scalable platforms, integration skills and best-of-breed components. Unisys [e-@ction](#) Multimedia Value-Added Services (MVAS) is a solution set that encompasses voice/fax messaging, universal messaging, calling services, Internet call waiting, and other enhanced messaging services. This announcement is significant since Unisys supports more than 100 million voice/fax mailboxes for over 100 MVAS clients in 40 countries. As part of the Unisys Natural Language Speech Solutions, the company provides professional services to design, develop and deploy spoken language applications to more than 50,000 customers.

#### **3.3.1.20 Vocalis Group, plc**

Vocalis Inc., formed in 1993, is the other major UK speech firm, besides SRC. Vocalis (Cambridge UK) bases its offering on its own advanced speech recognition technology sold as the Vocalis SpeechWare product. SpeechWare is now in Version 6.4 and has been used in Call Centers, by developers to create speech applications, and as part of a complete IVR solution. SpeechWare is platform independent and uses Unix operating systems. Through its Vocalis Solutions services it has developed solutions for businesses in: financials services, telecommunications and enterprise call centers.

Vocalis 'E-mail by Phone Solution' is an innovative service which allows your company employees to deal with their email over the 'phone while on the move, without even needing access to their computer! Imagine the efficiency and the time saved by allowing your employees to check their email from anywhere, without you having to invest in any additional equipment such as laptops or WAP enabled mobile phones. Vocalis 'E-mail by Phone Solution' is the link between any telephone and your email server. It allows employees to be notified by SMS when messages arrive and then make a phone call and hear their email read to them, they can forward the messages or email other contacts from their contact list. It's almost like providing them all with a personal secretary!

Vocalis announced a partnership agreement with the American Alliance of Service Providers, an organization for ISPs to combine efforts and gain increased purchasing power for additional products and services. Vocalis also announced additional ISPs who will add Vocalis SpeechMail to their Internet services, including Destiny PC, Clear Connections, Fastex, Village Online, Alti.Net, HiPlains.Net, TexasFamily.Net, konx.net, EON LLC, Quik Internet and Windmill.Net. Vocalis SpeechMail allows telephone access to existing e-mail accounts. Users send and retrieve e-mail messages using a combination of advanced speech recognition and text-to-speech technologies to manage email without a computer or direct Internet access. By eliminating the need for an expensive wireless digital assistant or computer to access the Internet, Vocalis SpeechMail offers an inexpensive, reliable solution to keep up-to-date while away from the home or office.

AASP member ISPs can sign up now to offer Vocalis SpeechMail to their customer base. SpeechMail for the other ISPs will go live in September.

### Key Developments

- Vocalis achieved sales from both Telenor Mobil, a Norway telecom and IT firm and Telefonos de Mexico (Telmex), Mexico's main telco, for Vocalis' Speechnel telephony platform. The new contracts are worth £395,000 and £260,000 respectively. Speechnel applications are based on SpeechWare, Vocalis speech recognition technology.
- Vocalis is now partnering with S2 Systems International, a global provider of business solutions to financial services firms. Development of home banking and other customer inquiry systems will be the goal of the partnership. S2 Systems is headquartered in Dallas, Texas.

### **3.4 New Developments Among Speech Application Vendors**

The intersection of firms involved in telephony and speech technology involves a diverse set of speech application vendors. This set comprises a large number of firms and diverse types of firms running the gamut of ASPs, switch makers, network infrastructure players, telephony application firms, and device manufacturers. The value chain analysis, provided above, provided important perspectives on these firms and their place in this arena.

This section focuses on firms who create platform solutions for specific telephony applications. In addition, they must include speech technology as a critical component to their solution. These application firms license speech technology from one or more of the speech technology software firms profiled in the section proceeding. These firms are successful at creating workable solutions in a variety of telephony applications. The most important have been identified as:

- Voice Activated Dialing;
- Mobile Handsets and Devices;
- Unified Messaging;
- Directory Assistance;
- Enterprise/Automated Attendant;
- Call Center;
- Personal Assistant;
- Voice Portals; and
- Telematics.

The firms chosen for this section are considered “best-of-breed.” The criteria used in the selection process was intuitive. In large measure, the essence of selection came about because these firms come up again and again in discussions with industry leaders, in industry forums and in the media. As companies, some of the issues common to these best-of-breed firms are that they are:

- Mentioned by key industry players in their 2001 speech-technology related activities;
- Key partners or customers of the speech technology software firms;
- Noted for excellence or received awards;
- Noted for having initiated or completed major new telephony speech service installations; and
- Important for having incorporated speech technology into a major product line(s).

### **3.4.1 Voice Activated Dialing**

The PELORUS Group report, *Wireless Speech Recognition*, profiles the key offerings in the wireless telephony space related to voice dialing. The most recent development has been the Cingular Wireless launch of both voice dialing and voice portals, supported



by BeVocal (see Voice Portals below). There are a number application providers providing either network-based or device-based voice activated dialing (VAD), including APEX Voice Communications, Aspect, Comverse, InterVoice-Brite, Motorola, Preferred Voice, and Pulsar. We profile Comverse because of its partnership with Nuance; Preferred Voice because of its success in selling voice dialing; and Toshiba because of their work with the major speech technology firm SpeechWorks.

### **3.4.1.1 Comverse**

Comverse has a voice activated dialing solution for network service providers. Comverse VAD is able to recognize a wide range of verbal commands, including terms such as call, dial, get, look up and phone. Users can dial from their personal address book, such as "Call Chris on her mobile phone" or "Call Jim Smith at work." The system also intelligently suggests alternative numbers when no one is available at the dialed number. System address books or even global phone books, such as yellow pages, can be utilized with Comverse VAD. Comverse VAD also works with its voice portal solution.

Comverse has an agreement with Nuance, which covers the entire Comverse voice-controlled enhanced services systems portfolio. This includes its [Tel@Go](#) Voice Portal. The Tel@GO voice portal provides natural language understanding capability, core voice portal applications, user personalization, and network integration, management and operations, all delivered on Comverse's carrier-grade platform. Nuance 7.0 provides a scalable, highly accurate speech recognition server capable of handling a very large vocabulary that is also ideal for use in noisy environments. In a command like "Check my stock portfolio," Tel@GO voice portal and Nuance 7.0 work together to recognize, understand, and act upon a user's spoken word in the context it is given while simultaneously incorporating the user's personal preferences

In addition to the Tel@GO voice portal solution, Comverse has integrated Nuance speech recognition software into the Progeny voice application hosting system designed for enterprise solutions. Progeny is a network-based solution that allows service

providers to host enterprise portals for corporate customers on a shared service platform.

#### **3.4.1.2 Preferred Voice**

Preferred Voice was mentioned for its considerable leadership in voice activated dialing at the Excellence in Marketing Conference held by the Rural Cellular Associations conference in 2001. Gold State Cellular provided a testimony of success using Preferred Voice VAD, in which it was able to achieve a 67 percent take rate across its subscriber base over a fourth month period. Preferred voice has signed contracts with some 45 telecom carriers for its VAD solutions. These carriers represent subscriber bases totaling 5.5 million lines in 36 states. So far, the number of VAD systems that have been implemented cover 820,000 customers with 15 carriers. Preferred Voice's voice dialing service is called Safety Talk. It includes other enhanced voice services including voice mail navigation, voice accessed content, voice calling cards, voice DA, and EmmaPR, which is a voice telephone receptionist.

#### **3.4.1.3 Toshiba**

Toshiba has integrated the SpeechWorks speech-enabling software into its Strategy Enterprise Server (ES). Toshiba has aimed at creating at enhancing its network server line, called Strategy Enterprise Server, to add voice-activation features to its messaging system. The first product to evolve from this partnership is an speech recognition auto-attendant that can be combined with other Strategy ES applications, such as voice mail and/or IVR. Toshiba's Strategy Enterprise Server and the SpeechWorks automatic speech recognition engine provide a most natural method of connecting callers to their destination. The result is cost-effective speech driven applications. Users can use their own voice to navigate menus and quickly retrieve information.

The Toshiba Telecommunication Systems Division markets the Strata DK family of digital business telephone systems, the Strata CS communication server, Strategy voice processing systems, and Interactive Voice Response systems.

### 3.4.2 **Mobile Handsets and Devices**

Voice dialing has been and continues to be an option on some of the telephone handsets that are sold by the wireless and wireline carriers. Some of the handset vendors use QUALCOMM's (San Diego, California) technology. QUALCOMM offers PureVoice VR2.1, which is a handset chip-based processor that enables users to access phone functions and dialing by voice. It supports a 40-name directory. Another vendor, Sensory, Inc. (Sunnyvale, California), has a processor that adds voice dialing to a number of devices including mobile phones, hands-free automobile kits, and PDAs. And Telespree, using Nuance's leading speech recognition technology, has developed a unique "keypadless" one button (on-off) handset. The completely voice-driven telephone is a prepaid mobile phone with a replaceable, combination battery pack and replenishment chip.

Speech technology is found on more than just phones. In fact, speech dictation and other forms of interacting with PCs were some of the earliest applications. Lernout & Hauspie Speech has always been a leader in embedded speech technology in computing products, but due to bankruptcy proceedings has sold many of its assets to ScanSoft. L&H technology will continue to be used and expanded into many forms of mobile devices. Its Pdsay software provides users with eyes-free and hands-free access to PocketPC information. Through natural spoken commands users can navigate through inboxes, contact databases, and appointment calendars -- hearing emails and appointments read out loud via L&H's TTS technology. Listed at \$39.99, the software works with a Compaq iPAQ 3630 Pocket PC or higher and the Windows CE 3.0.9348 or greater.

### **3.4.2.1 Andrea Electronics**

Andrea Electronics Corporation recently received a Peak Performance Award from Speech Technology Magazine. The award program is designed to recognize the products, services or innovations of speech technology companies whose efforts have propelled the industry to the forefront of technological innovations in communications. In addition, Andrea Electronics's technology was highlighted in demonstrations as part of the Delphi Automotive Systems Concept car at recent automotive shows. Andrea Electronics' telematics product, DA-310 AutoArray, is a digital microphone system that is designed for use with in-vehicle voice-activated applications. Delphi Automotive Systems has developed the latest in automotive telematics applications, including its COMMUNIPORT Mobile Multimedia Systems. Andrea's DA-310 AutoArray microphone will be used to demonstrate command and control and voice-activated cellular phone functions for the COMMUNIPORT system.

Andrea Electronics was formed in 1934 and its first IPO was in 1960. Its embedded speech technology can be found in all kinds of mobile devices, including mobile phones, PDAs, laptops, pagers and telematics control units.

### **3.4.2.2 ART**

Advanced Recognition Technologies (ART) has integrated its technology with that of Persay to offer solutions for m-commerce. Persay is a Comverse Technology spin-off and a leader in voice verification solutions. The solution will combine ART's speech recognition technology for mobile devices with the additional Persay voice verification engine to provide greater security for mobile handsets and various wireless devices.

Motorola's Semiconductor Products Sector uses ART's speech and handwriting recognition software on Motorola's new microprocessor line. In addition, Motorola and ART collaborate in the co-marketing of ART-ported products. ART's leading smARTspeak NG and simpliWrite technologies are included in Motorola's processors,

including the 2/2.5G and 3G Innovative Convergence Platform and DragonBall applications processor families.

Advanced Recognition Technologies products are also used by Mitsubishi Electric Telecom Europe (METE), the sixth largest manufacturer of handsets worldwide. The global agreement between the two companies involves integration of ART's smARTspeak and smARTspeak CS speech recognition solutions into the Eclipse and Sirius models of METE's Trium mobile phone family. ART's smARTspeak software enables a mobile user to call a person from the address book of their phone by simply saying the person's name. smARTspeak is a well proven noise immune product, which allows for true hands-free operation.

A further enhancement to smARTspeak, the CS version enables continuous speech digit dialing. Speaker-dependent, it allows users to dial a number by simply saying the string of numbers - with no pause between digits. ART's smARTspeak CS is one of the few language-independent continuous digit dialing technology available in the market today.

ART was established in 1990 and its embedded software (smARTspeak NG and smARTcar) are deployed in products lines of companies like: Acer, Agere Systems, Casio, Franklin, Hitachi, Konka, LGE, Maxon, Microsoft, Motorola, NeoPoint, Panasonic, Philips, Samsung and Siemens.

### **3.4.3 Unified Messaging**

The PELORUS Group's Unified Communications for the Service Markets report conducts a thorough treatment of current unified messaging and unified communications offerings and players. The report identified those that have incorporated speech technology, especially TTS, into their offerings. The carriers that were profiled in the report are: SBC, Bell Canada, BellSouth and Cincinnati Bell. Platform/Application firms included: BridgeCom, Cisco, Comverse, Ericsson, InterVoice-Brite, Lucent, Mediagate, Virtualplus, Tornado, Unisys, and Vocal Link. The report also

covered service providers and ASPs that are using speech technology, which include: DeltaThree, eTouch, Linx, Verso, and ShareMedia.

There is overlap between UM and the other application areas profiled in this chapter. Unified messaging is also linked with services such as voice portals and call centers. The topic of unified messaging and speech was also explored in The PELORUS Group's recent report, *The Future of Contact Centers*, and it was revealed there that certain call center vendors are involved with unified messaging solutions, including: Interactive Intelligence, Mitel, Perimeter Technology, Siemens, and Toshiba.

Most UM is now offered in enterprise settings. But unified messaging for mobile consumers is also available and some of these incorporate voice-activated interfaces. For example, the unified messaging company, eVoice (based in Menlo Park, California) recently extended their UM service to mobile phones. As a voice portal ASP they offer wireless carriers the capability for Internet plus voice-enabled messaging. The basic eVoice service alerts the subscriber as soon as someone leaves them a message. The alert goes to the subscriber's email, pager, and mobile phone accounts. Users can check their voice mail from anywhere using a toll-free number.

One must not ignore the basic voice mail services though. Philips Speech Processing, for one, is convinced that one critical application of speech technologies will be voice-activated access to basic voice mail. They say they are seeing thousands of ports added by their customers in the move toward making voice mail mobile-friendly. Philips is working with various ASPs to help some 15 different carriers upgrade their voice mail systems with speech technology. One of Philips' partners, Comverse, is working with them to create voice-enabled voice mail applications.

#### **3.4.3.1 InterVoice-Brite**

InterVoice-Brite's (iVB's) UM product has been relaunched as part of its Omvia initiative. Omvia Messaging is part of iVB Network Solutions' (InterVoice-Brite) Omvia family of scaleable, data-centric applications. Omvia solutions allow users to seamlessly

navigate among multiple value-added network applications in a single session. Employing state-of-the-art speech and visual (Web and WAP) interfaces, Omvia solutions enrich the experience of users.

Speech technology is being infused into other Omvia applications as well, including secure m-commerce, payment services (prepaid and postpaid) and content (information services). InterVoice-Brite offers these capabilities as part of its CeASP (Communications and eBusiness Application Service Provider) solution set. The e-business application is speech enabled and offers full hands-free operation, particularly suited to the wireless Web applications. The solution uses a wide selection of components from the InterVoice-Brite advanced technology suite, including text-to-speech, WAP, Lightweight Directory Access Protocol and market-leading hardware components. The TTS solutions are derived from iVB's TTS Server. The InterVoice-Brite TTS Server can be equipped with a number of TTS engines, Lernout & Hauspie's Realspeak, Elan Speech Cube or Telia Promotor engines may be used. iVB's voice recognition capabilities are based on Nuance's voice recognition algorithms.

### **3.4.3.2 Mitel**

Actually Mitel Networks offers a set of speech technology solutions, rather than just unified messaging. Mitel is involved in call centers, auto attendants and UM. Mitel Networks (formerly the Mitel Communications Systems division of Mitel Corp.) provides IP communications platforms, applications and solutions in the areas of speech recognition, wireless mobility, unified messaging, and contact center solutions. The company has regional operations located in Herndon, Virginia (US Sales), Caldicot, Wales (European headquarters); and Singapore (Asia-Pacific operations). Mitel Networks operates 71 regional facilities in the US, Canada, the UK, Europe, and the Far East. Manufacturing facilities are located in Canada and the UK.

The Mitel Networks 6500 series of speech-enabled applications are focused on telephony solutions. The applications are designed to provide a number of benefits: free

users from having to remember a myriad of telephone numbers and pass codes; eliminating the frustration of dialing extensions by spelling last names; and allowing users to make calls and access sophisticated telephony features simply by telling the system what to do.

Mitel Networks 6500 Speech Enabled Unified Messaging solution is a unified messaging solution that features natural speech as a user interface for accessing, manipulating, navigating and responding to all messages in a single unified inbox. For the call center market, Mitel Networks also offers its Mitel 6500 Speech-Enabled Attendant. This auto attendant easily integrates with Mitel Networks enterprise IP telephony and PBX platforms to support a wide range of advanced features such as Lightweight Directory Access Protocol (LDAP) integration, advanced text-to-speech, and recognition of voice signatures.

### **3.4.3.3 Nortel**

Major UM vendors, like Nortel, have integrated speech recognition into their systems. The Nortel UM system, called CallPilot, has built-in speech recognition that provides voice interfaces to voice mail and fax messages. The capability is now being expanded for voice access to email as well. CallPilot is Nortel's unified messaging tool that brings together voicemail, email and faxes to create a personal communications management system that can be easily used through voice commands from virtually anywhere. CallPilot is an application common to a number of Nortel systems, including the Meridian 1 PBX, the Succession Communications Server for Enterprise, M6500 (Europe, Africa and Middle East only), Passport and Business Communications Manager platforms.

Nortel works with SpeechWorks, Fonix and Nuance for their speech technology. Nortel Networks and SpeechWorks have worked together to combine the benefits of Nortel Networks OSCAR (Open Signal Computing and Analysis Resource) IVR System, with OpenSpeech Recognizer, SpeechWorks' next-generation speech recognition engine



optimized for VoiceXML environments. SpeechWorks OpenSpeech Recognizer provides an open standard system that facilitates easy integration with Nortel Networks in both the OSCAR proprietary environment, as well as in the upcoming Nortel Networks VoiceXML offering. Nortel Networks will be able to leverage OSR effectively and efficiently on both product lines.

Nortel has also signed an agreement with Fonix Corporation to use its TTS engine (called FFAST) for use in the Nortel OSCAR platform. Fonix will provide Nortel Networks with the voice processing engine for use in OSCAR IVR components. FFAST will provide Nortel customers with TTS in multiple languages.

#### **3.4.4 Directory Assistance**

The recent PELORUS Group report that relays information about speech technology in directory assistance is called Enhanced Directory Assistance. It profiles the EDA providers in detail and discusses their application of speech recognition. The report profiles for VoltDelta, varetis, ISx and Nortel all discuss speech technology implementations. The best-of-breed firms reported below are: ISx, varetis, and VoltDelta.

##### **3.4.4.1 ISx**

Information Services eXtended (ISx, Inc.) is a DA services platform provider, headquartered in Fort Lauderdale, FL. ISx was launched as a separate company from IBM in January 2001. However, ISx has a twenty-five year history in directory assistance with its roots as an IBM and Bell Labs research project dating back to the 1960's.

The latest ISx Directory Assistance product is the new ISx4 platform, which was introduced in May 2001. The new platform has been expanded with speech recognition and Internet access for cost-cutting direct access to listing data. Speech recognition

through this ISx Call Automation product is designed to improve call center efficiency. It can handle the DA needs of a small call center or a multi-location call center with over 3,000 workstations.

ISx offers both audio information delivery and speech recognition. It provides integrated speech recognition for automating Directory Assistance through its Automated Operator Assistance Platform. The platform includes:

- Recording the inquiry through analysis by the speech recognition engine;
- Generation of the database inquiry; and
- Audio delivery of the listing information.

For requests with DA operator involvement, the operator listens to audio playback of the caller's request with the workstation already displaying possible listings. When only the locality is recognized, the search screen is delivered with a locality pre-populated, along with the recorded inquiry. The base application, Voice Store and Forward, prompts the caller automatically, and the request is recorded. The silences are trimmed for playback to the operator. Prompts and systems options can be customized for each specific market.

The ISx Intelligent Audio Subsystem automates the delivery of requested listings. After the operator selects the correct listing, the information is released for audio delivery to the caller. SMS delivery of the listing data to a wireless handset is also an option.

#### **3.4.4.2 varetis**

According to varetis officials, they are working with Nuance, IBM and Philips in order to incorporate speech technology software into varetis automated DA solutions. Formerly pc-plus, the varetis Group offers directory and operator services solutions for the telecommunications industry. varetis has been a major player in the international directory assistance systems market for a number of years and is gaining recognition in the US market for DA products and integrated operator services solutions. Over 240

call centers and 18,000 DA operators worldwide work with solutions from varetis. varetis offers both a standard solution for national and international Directory Assistance as well as customized solutions.

The parent company, varetis AG, is headquartered in Munich, Germany and employs 280 people. The American arm, varetis COMMUNICATIONS L.P., is based out of Elmwood Park, NJ. Other locations include Brazil, Hungary, Switzerland, United Kingdom, and France. varetis has been in business since 1983.

The key NDA offering from varetis is the National Directory Inquiry System (NDIS). NDIS is an integrated, UNIX-based search engine, specifically designed for Directory Assistance applications. Its search capabilities include nationwide, phonetic, and expanded locality searches. NDIS supports all standard directory searches (residential, business, or government). Based on its advanced architecture, NDIS can accommodate a multitude of fields in variable lengths. Varetis has been using Phillips Speech Processing for solutions that partially or completely automate DA Applications.

### **3.4.4.3 VoltaDelta**

VoltDelta's SpeechExpress is an open speech platform that automates directory assistance by using speech recognition, store and forward voice and voice portal technologies. SpeechExpress offers its customers various degrees of automation including:

- Full -- The SpeechExpress system handles everything from prompting requests from callers to call completion;
- Partial -- The SpeechExpress system handles prompting, recording, and analyzing caller requests, yet offers intelligent call routing to operators when a caller so chooses; and
- Store and Forward -- The SpeechExpress system prompts, records, and forwards caller requests to operators for call completion.

Designed to automate directory and information services, SpeechExpress offers a variety of custom applications from basic audiotext to advanced applications, such as the following:

- Text-to-Speech Automation;
- Directory Inquiries Automation;
- Automated Reverse Directory;
- Voice-Activated Dialing;
- Fax Messaging;
- ACD;
- Voice over IP (VoIP).

Automated information services from VoltDelta include sports, weather, stock quotes, movies, and a host of others. Customers of VoltDelta include AT&T, Cincinnati Bell, Verizon, SBC, SNET, Cable & Wireless, KPN, Go2 Systems, and Etak Inc.

### **3.4.5 Enterprise/Auto Attendants**

It is difficult to summarize the enterprise market, since the vendors serving this marketplace offer so many different kinds of automation solutions. In addition, solutions that serve the basic enterprise also can be applied in the call center arena as well. The key solution sets in this market relate to enterprise auto attendants and call center IVR enhancements.

Most the vendors have the same basic auto attendant and automated routing feature sets. Specific auto attendant firms are: Phonetic Systems, Sound Advantage, Lyrix, Parlance, Locus Dialogue, Mitel, ONEder Technologies, and Artifsoft. Also refer back to the profile of both InfoSpace and LocusDialog in the speech technology software vendors above, since they both serve the enterprise market. Part of LocusDialog was sold to InfoSpace and the part that remains exclusively focuses on the enterprise space. Mitel and Toshiba are also in this space, but are profiled under unified messaging and call center sections, respectively.

### **3.4.5.1 Interactive Intelligence**

Interactive Intelligence Inc. is a global developer of multi-channel customer interaction management software that allows e-Businesses, enterprises, contact centers and service providers to automate virtually every aspect of their business communications. The company was founded in 1994 and, today, has a worldwide customer base of more than 800 companies. Interactive Intelligence's global headquarters are located in Indianapolis, Indiana with European headquarters in France and Asia Pacific headquarters in Japan. The company employs more than 400 people located in offices around the world.

Interactive Intelligence's Interaction Center Platform is the basis for its suite of Windows 2000-based products designed for call centers, which it calls interaction centers. The platform integrates and aims to replace proprietary communication devices such as PBXs, ACDs, IVRs, voice mail systems, fax servers, Web gateways and computer telephony integration middleware systems with an integrated system. Unlike traditional telecom solutions, the Interaction Center Platform offers a unified architecture that virtually eliminates the need for costly step-by-step integration. This architecture reduces start-up and maintenance costs, simplifies administration and increases ease of customization.

Interactive Intelligence offers a single multi-media solution for contact centers that is an alternative to integrating various forms of Web-based and legacy voice technology. This includes use of speech-enabled interfaces. The most recent launch from Interactive Intelligence is its Mobilite 1.0, a wireless application gateway product designed for enterprises. Mobilite helps enterprises deploy business applications for PDAs that have wireless connectivity. Mobilite supports a company's effort to support mobile workforces and multiple types of mobile devices. Mobilite supports a variety of networks (including 802.11 and GPRS) and includes text-to-speech and speech recognition capabilities.

### **3.4.5.2 InterVoice-Brite**

The iVB Enterprise Solutions Division has more than fifteen years experience in the development and deployment of speech and self-service solutions for the enterprise market. Its systems serve millions of its enterprises' customers and institutions' patrons worldwide. The company has implemented speech-driven applications for travel and flight information, bank-by-phone, stock quotes and trades, hospital and pharmaceutical call directors, customer service help desks, and ordering applications.

The company's SpeechAccess solution combines the customer self-service and application development products from InterVoice-Brite with the speech recognition, speaker verification and text-to-speech technologies of companies like SpeechWorks, Nuance and L&H. Using the capabilities of SpeechAccess, InterVoice-Brite can design and deliver speech-enabled applications and voice portals for many types of companies.

InVision is application development software module of iVB Enterprise Solutions that enables customers to create, document, modify and test advanced call flow scripts. It allows businesses to prototype and deploy multimedia applications such as voice portals with speech recognition, text-to-speech applications and Web interfaces. InVision capabilities can be further expanded with vendor-specific speech toolkits featuring SpeechWorks, Nuance and L&H RealSpeak products. An application developed with InVision can connect with virtually any communication device, through any switch, PBX system and over any IP network to any telephony host or back-end database. InVision also supports creation of applications featuring advanced functions including speech recognition, text-to-speech and faxing.

### **3.4.6 Call Center/IVR**

In the recent PELORUS Group report, called the Future of Contact Centers, the issue of speech technology among call center vendors was thoroughly discussed. The

report profiles call center platform vendors and details their offerings. The firms identified as already upgraded their solution set with speech technology are: Aspect, Avaya, Interactive Intelligence, Mitel, Nortel, Siemens and Toshiba.

### **3.4.6.1 Aspect**

Through acquisitions and partnerships Aspect has evolved into a company that provides comprehensive contact center and CRM solutions. Aspect's products enable companies to accumulate and deliver customer information when and where it is needed. Additionally, Aspect has provided the GUI-based development tools that enable companies to implement and tailor CRM strategies quickly and efficiently.

Aspect Customer Self-Service (CSS) is a completely software-based IVR solution that supports open systems. Companies purchase the software from Aspect and separately purchase the hardware and operating systems from infrastructure vendors such as IBM, Compaq, and others. Aspect CSS provides customers with access to any business information via speech recognition, fax, phone, or the Web. The application controls corporate media resources while integrating disparate information sources, to provide customers with personalized services. Aspect CSS can be implemented in a stand-alone environment with any ACD, or as part of an integrated CRM solution in conjunction with the Aspect Enterprise Contact Server (ECS).

Aspect CSS and Aspect ECS together support call center applications. One important feature is if a call needs to be elevated from self-service to live-service, there is no loss of customer data, as the call history collected by CSS is passed along to the agent via the Enterprise Contact Server. The ECS product comes with a GUI interface for enterprise staff to use in developing workflows.

In June 2001, Aspect began launching the first in a series of planned wireless applications. Aspect Mobile Banking v1.0 provides live and self-service applications to customers using Palm devices and cell phones. The solution can be integrated with other transaction channels such as home banking. Mobile Banking lets consumers

access a range of services for their accounts, such as accessing account balances, bill payment and funds transfer in real time through their mobile devices.

#### **3.4.6.2 Avaya**

In October 2000, Lucent Technologies officially spun-off its Enterprise Networks Group, which provided PBX, SYSTIMAX structured cabling and LAN-based data products. The newly formed Avaya Inc. became the provider of what was Lucent's enterprise contact center solutions including CRM software and PBX switches. Avaya's Enterprise Communications business provides solutions for a broad range of business communications needs through several core product-group organizations:

- Customer Relationship Management (CRM) includes traditional inbound and proactive outbound call centers, multi-channel contact centers, voice and speech self-service applications. Its aim is automation of customer-driven business processes across the enterprise and unleashed business intelligence.
- Unified Communications includes voice messaging, unified messaging, conferencing and "any-media" speech and wireless enabled usage.
- Avaya's Interaction Management is comprised of the Avaya Interactive Voice Response product for automated self-service and Avaya Natural Language Speech Recognition product for advanced speech applications.

Avaya is working to leverage its existing leadership in the traditional call center market, earned through its DEFINITY and CenterVu product families, in order to expand its market share by delivering solutions in a host of new areas. They include multichannel contact centers, proactive inbound/outbound centers, distributed IP contact centers, speech and advanced voice response applications, Web and email management and self-service, and commitment management business automation.

Avaya intends to accomplish its goals through integration of third-party solutions as well as in-house product development. Avaya is in the process of building major CRM partnerships. These include agreements with Blue Pumpkin for workforce



management, Nuance and SpeechWorks for voice applications, Cognos for business intelligence, Siebel for applications, and others. Sales distribution and global services are provided through IBM and Unisys.

The product most relevant to this report is Avaya Interactive Voice Response. It is a scalable self-service voice response and speech platform that is part of the Avaya Interaction Management CRM solution suite. It has been installed in over 22,000 systems in over 55 countries. It utilizes Compact PCI platforms with redundancy and hot-swap capabilities and current supports up to 192 ports per platform, which will expand to 512 in releases in 2002. The v8.0 release last year opened up the platform to make integration with external applications much easier through Java and XML, with interfaces to WAP, HTML, XML, Siebel, MS SQL2000, Nuance SpeechObjects, and SpeechWorks. Scripting tools are GUI-based, open interfaces to both CRM Interaction Management and CRM Commitment Management that allow IVR and speech processing integrated where necessary. In 2002, Avaya will introduce on-board VoiceXML for both the current v8.0 platform and new 2002 platforms.

Working with both Nuance and SpeechWorks (Avaya resells both) the company has been aggressively deploying new speech-based applications to many customers, totaling over 30 within the last 12 months. It utilizes a client/server speech integration design that allows multiple IVRs and multiple separate speech servers to work together in various "shared resource" or redundant combinations. Avaya provides a direct linkage through Avaya Interaction Center to the rest of the contact center, allowing routing based on IVR data, customer-specific IVR and speech script. The IVR and speech platforms integrate with other ACDs, including Nortel, Aspect, Rockwell and others, and Avaya sells IVR, both stand-alone and in conjunction with Avaya's end-to-end contact center portfolio. Avaya has been significantly growing its IVR partnership program in recent quarters and now has over 18 certified partners covering application development, vertical industry specialization, Web integration expertise and human factors consultation.

### **3.4.6.3 Nortel**

Nortel's Speech Processing Platform provides multi-function speech processing capabilities in an open client environment. Known as OSCAR (Open Signal Computing and Analysis Resource), this platform incorporates Large Vocabulary Recognition (LVR), Natural Language Understanding (NLU), Text-to-Speech (TTS), and Speaker Verification technologies. The OSCAR architecture is designed to support advanced speech-processing applications in an open and scalable environment.

In October 2001, Nortel Networks and SpeechWorks announced that they would combine Nortel Networks OSCAR with OpenSpeech Recognizer, SpeechWorks' speech recognition engine that is optimized for VoiceXML environments. Combined also would be the SpeechWorks Speechify TTS engine and SpeechWorks SpeechSecure speaker verification module. By combining Nortel Networks speech technology deployment experience with SpeechWorks user interface expertise and speech solutions, the companies expect to deliver a joint solution capable of providing user-friendly natural language speech applications such as order fulfillment, flight information, banking transactions and other services. This solution, scheduled to be available in early 2002, is expected to offer tailored solutions for the financial, telecommunications, healthcare, travel and insurance industries.

Nortel Networks' business strategy includes a recent restructuring of the company into key focus areas. The Customer Contact and Voice Portal Solutions line of business, now organized under the Metro-to Enterprise Networks organization, continues to be a key component in the company's commitment to the enterprise business. The Customer Contact and Voice Portal Solutions group has identified several major drivers in the contact center market, one of which is the economic downturn. In evaluating the contact center market within this economic environment, it became evident to Nortel that its contact center customers and voice portal customers had several primary interests. One was to reduce customer churn and strengthen their customers' loyalty through implementation of CRM strategies.

Another primary driver for Nortel contact center customers is reduction of operating costs. Nortel is examining how it can best reduce these costs for its customers, and one of the answers is provision of technologies that automate contact center tasks in a user-friendly way. Nortel is focusing on development of self-service solutions, provided both through the Internet and through advanced speech services. Nortel IVR and advanced speech-enabled solutions are developed as modular components of a suite of solutions that can be added to a system as needed.

In October 2001, Nortel Networks and IBM announced plans to deliver a suite of hardware, software and services to provide customers with open-standards converged voice and data networks. Nortel Networks and IBM plan to incorporate IBM WebSphere e-business infrastructure software and DB2 database software into Nortel Networks Symposium Contact Center solutions portfolio. These activities will include open-standards based, integrated VoIP products from Nortel Networks with speech-enabled interfaces.

#### **3.4.6.4 Toshiba**

In March 2000, Toshiba entered the communications server-based telephone market with the introduction of the Toshiba Strata CS. Strata CS is a software-driven business communications system that provides call management from the desktop PC through a Strata server with PBX and voice mail functionality.

Several Strata CS server models are available to accommodate the needs of both small and large organizations. Systems start at 4 trunks and 8 stations and the largest system has a maximum of 96 trunks and 264 stations. Each Strata CS Server system contains the same features and functionality, using the Strata CS application software.

Strata CS provides both IP and circuit switched communications. It is compatible with both IP phones and traditional telephones (digital and analog). Users can connect an IP telephone locally to the system through a LAN or can connect an IP telephone remotely to the system through the Internet or intranet. Strata CS also provides

industry specific, pre-packaged IVR applications using speech recognition and text-to-speech.

The other important Toshiba product is Strategy ES, which provides a platform for voice processing applications, including IVR, voice and fax mail, unified messaging with text and speech. ES has a modular, scaleable design built on the Microsoft Windows NT operating system. It is available for implementation as a custom turnkey or with solution templates. It is scalable from four to 96 ports. SpeechWorks International provides the speech activation technology.

The functionality of the ES is provided through a family of product applications called Application Suites and Feature Groups. Application Suites are Interaction Voice Response and Messaging Voice Processing (MVP), which can function with each other or with any combination of Feature Groups. Feature Groups currently available for IVR include Fax Server, Automated Attendant with Speech Recognition, Host Interface and Text-to-Speech. For MVP, the Fax Server Feature Group is available as an option.

### **3.4.7 Personal Assistants**

The PELORUS Group will be announcing publication of a new report on Unified Communications this spring, 2002. The previous report, Unified Communications for the Service Markets, conducts a thorough treatment of current unified messaging and unified communications offerings and players. The report identified those that have incorporated speech technology, including virtual personal assistants.

#### **3.4.7.1 Webley**

Webley Systems, Inc. is one of the leaders in the fast-emerging personal assistant services marketplace. Webley Systems is working on creating systems for voice-activated unified messaging and the whole gamut of unified communications. Webley offers two core virtual personal assistant products: the Personal Assistant and Corporate

Assistant. Webley Personal Assistant is a virtual communications manager with Web access that can be used to make calls, screen callers, and take messages. Webley also reads email over the phone (text-to-voice) and collects and forwards faxes.

Corporate Webley Assistant is a single number that acts as an auto attendant front-end for several individual mailboxes. When calling the corporate Webley Assistant number, the user is prompted to speak the name of the individual or the department they are trying to contact. The call is then forwarded to the individual or their mailbox. Corporate mailbox owners are also given their own direct access number. In this way, they can receive faxes directly in their mailbox and gain faster, more direct access to the system.

Using one phone number, a user can receive calls, faxes and voice mail. Webley offers features such as call-forwarding, access to an address book, conference calling and email by phone, as well as a virtual attendant that operates through voice recognition. Webley's most popular package costs \$15 per month, plus additional usage fees: 6 cents per minute for inbound calls and 5 cents per minute for outbound long distance calls.

Webley also offers its CommuniKate Unified Communications (UC) product. Using a highly scalable IP-based operating platform and world-class speech technology, UC simplifies communications by offering a single phone number for handling inbound IP/PSTN calls, voice mail, e-mail, and faxes. Users can update contact lists, manage conference calls, schedule reminders, and set personal preferences using either their phone single-number access or personal Web page. Another component of CommuniKate is Conferencing and Collaboration. Webley offers a state-of-the-art integrated or standalone Conferencing product that supports multi-party scheduled and on-demand conferences and information collaboration. Users can control these calls by telephone, issuing DTMF or speech commands and by an easy-to-use Web interface that allows for conference call control as well as information sharing and collaboration.

### **3.4.7.2 Wildfire**

Wildfire, which was acquired by Orange plc Group, had created one of the first speech-activated personal assistants. Wildfire has changed its strategy so that it now primarily sells through wireless carriers due to the proliferation of wireless users. The company also has plans to work with ISPs and wireline carriers.

Wildfire uses speech recognition to manage communications functions in its services, such as voice dialing, voice mail, call routing, email and fax through a single voice-controlled interface and a single personality. Subscribers can choose among Wildfire's more advanced features to build their own customized Wildfire Assistant. Network Wildfire runs on industry-standard Compaq ProLiant Pentium-based servers. Wildfire sells the wireless carrier the software, which runs on a Compaq ProLiant server platform. Then Wildfire loads the software onto the server and integrates it into the wireless carrier's network.

Basic personal assistant features include:

- Voice-controlled Voice Mail;
- Basic Dialing; and
- Call Screening.

And enhanced services include:

- Unified Fax Messaging: storing them, forwarding them to wherever;
- Contacts Address book management with name dialing;
- Follow Me Routing;
- Call-On-Call Intelligent call waiting;
- Advanced Messaging;
- Conference Calling;
- Reminders;
- Unified Email Messaging;
- Virtual Hallway: knowing who's around with "buddy lists;" and
- Web Browser Interface.

Users can sort and screen messages, make calls from their contact list, and do other functions using speech recognition or the touch-tone phone. Wildfire Communications has built a combined voice dialing and voice mail application that seems attractive. In Wildfire Personal Assistant, the basic service allows the user voice dialing and voice-activated voice mail.

The company's vision is to move beyond telecom only to overall life services management, covering all of a person's communications. The Wildfire assistant will be more proactive and do things such as make and confirm appointments, retrieve content, and be a part of the transactions rather than just handing the user off. The system currently supports four languages, including North American English, British English, French, and Italian.

As part of Orange, Wildfire has expanded into international services, including Bouyge (France), which renamed Wildfire as Lucy and put a visual face on it, as well as Blu (Italy), which branded the Wildfire product as Memory.

### **3.4.8 Voice Portals**

The voice portal market is undergoing dramatic change, as portal providers move away from direct provision of the end-user, consumer-oriented portal to partnerships and service agreements with the carriers, especially wireless carriers. The PELORUS Group report, Wireless Speech Recognition, profiles the various wireless Web and voice portal players and offerings.

#### **3.4.8.1 BeVocal**

In the largest wireless carrier deal to date, the portal ASP BeVocal signed up Cingular Wireless to use its "voicetone" software and hosting services to power Cingular's first voice portal product, known as Cingular "Voice Connect." The Santa Clara, CA-based company has struck similar deals with BellSouth and Qwest Wireless.

Voice Connect lets customers use voice commands to access information services such as stock quotes and sports. It also includes a voice dialing service, which lets customers speak a name or phone number to make a phone call as well.

The service is now available to some 7 million customers in nine Southeastern states including Georgia, North and South Carolina, Tennessee, Kentucky, Florida, Alabama, Mississippi, and Louisiana. Cingular will roll out the service to its other 14 million customers in the near future. Cingular Wireless customers can activate Voice Connect service by dialing \*8 and following the prompts. By activating Voice Connect, the first 60 days of service free. After that Cingular will charge you \$4.99 per month for its Voice Activated Dialing (VAD), \$4.99 per month for Voice Information Services (VIS) and \$7.99 a month for both features.

BeVocal's platform lets you listen to, delete and respond to email messages over the phone. The company's address book has storage and retrieval capability of up to 500 contact listings over the phone, and synchronizes with personal information managers such as Microsoft Outlook and Palm Organizer. BeVocal Notification lets you specify and schedule automatic alerts for time-sensitive information, while the Voice Authentication service wraps the whole bundle in a secure package through the use of voiceprints to verify the identity of a caller.

BeVocal has created a reusable applications platform called VocalSuites. One of reusable applications is the BeVocal VocalLocator application, which enables service providers to create portals where users can call in for information about nearby businesses as well as get directions to those locations. The VocalLocator solution directs people to hotels, restaurant chains, retailers, banks and other businesses.

### **3.4.8.2 HeyAnita**

HeyAnita was one of the early leaders to sign up a major wireless carrier for voice portals. It struck a deal with Korea Telecom, which offers HeyAnita services to its



customers. Two-year-old HeyAnita announced a third round of funding in late 2001 from Intel and other investors, bringing its total financing to date to \$40 million.

HeyAnita VoiceXML browser ranked the highest in several critical categories in browser industry tests. In CT Labs Testing Services Report measuring browser performance, HeyAnita received a perfect 10 points for both "Voice Performance" and "Error Detection", with CT Labs detecting only one error throughout the entire 600 call automated test. CT Labs aimed to investigate how easy it was to create a VoiceXML application from scratch and get it running on each of the four voice portals participating in the test: HeyAnita, Tellme, BeVocal and VoiceGenie. HeyAnita's Free Speech Platform allows programmers to use their programming language of choice and the Platform will transform that code into VoiceXML.

HeyAnita also partners with telephony platform firms. For example, Net2Phone and HeyAnita jointly market end-to-end voice hosting solutions to enterprises, call centers, broadband service providers and its established user base. Net2Phone provides the telephony infrastructure and hosting platform and HeyAnita builds voice applications, and provide support and professional services.

### **3.4.8.3 Tellme**

So far, Tellme has obtained the largest amount of financing among the voice portals with some \$238 million coming from a combination of AT&T and venture capital firms such as Kleiner Perkins Caufield & Byers, Benchmark Capital and Barksdale Group.

Tellme has the most comprehensive suite of services so far. The number of information categories available varies by service but Tellme has more than others, with a total of 14. Tellme is reviewed by the media as doing the best job so far of arranging a set of useful and in-depth services. One plus for Tellme is that it has a database of 450,000 restaurants nationwide. Plus it includes the famous Zagat restaurant reviews for the 20,000 restaurants covered by the Zagat guides. Tellme also has a service that helps you call a taxi from wherever you are.

AT&T offers the Tellme driven voice portal for its wireless #121 service. It is also offering voice portal services to its business customers to help them speech-enable their Web sites and call centers. To do this, AT&T is offering integration services from their outsourcing unit, using its existing 800 Service network and adding Tellme's technology. The resulting partnership will be able to offer ways to create voice-activation interfaces to customer Web sites and to build voice-activated response processes for customer call centers using speech technology.

### **3.4.9 Telematics**

Essential to providing telematics services is location-finding technology, most often GPS. A GPS receiver is embedded into the telematics device that is mounted into the dashboard of the car. Using GPS, the telematics service can identify the location of the vehicle. The location-based telematics creates a link between any vehicle and virtually any information or service provider in the world. Cars as well as commercial vehicles become links on the information highway, as they constantly exchange real-time information with the outside world, to assist vehicle drivers and occupants.

The PELORUS Group just published a new report on [Location-Based Services Markets](#), which covers the whole range of LBS services, including voice portals and telematics. Essential to these offerings are speech-enabled interfaces, which make use of them easier for the mobile customer.

#### **3.4.9.1 Visteon**

Visteon Corporation is a leading full-service supplier that delivers consumer-driven technology solutions to automotive manufacturers worldwide and through multiple channels within the global automotive aftermarket. Visteon has about 80,000 employees and a global delivery system of more than 130 technical, manufacturing, sales and service facilities located in 25 countries.

Currently, Visteon Voice Technology is available to consumers on several production vehicles. It allows drivers to use voice commands in the vehicle to adjust the temperature, tune the CD and radio, make phone calls with voice activated dialing capabilities such as, "Call dad at home."

To further expand its offerings, Visteon Corporation and SpeechWorks International have a joint licensing and development alliance with the goal of creating a vast array of voice-enabled products and applications for both the automotive and consumer markets. This partnership is expected to take the voice and telematics industry to the next level. Visteon and SpeechWorks are working together to bring to market new speech-enabled products and services for telematics, multimedia, information services, personal productivity, banking and finance, and other applications.

Under the terms of the agreement, SpeechWorks is licensing Visteon's C-REC technology, a continuous, speaker-independent, phonetic-based speech recognition engine, designed specifically for embedded devices. Together with Visteon, SpeechWorks will enhance C-REC capabilities with its own speech recognition technology and user interface design expertise. This will accelerate C-REC's deployment into vehicles and other devices and the development of distributed speech recognition (DSR) systems. These enhancements will be targeted at applications such as mobile telephony handsets, personal digital assistants (PDAs), and cable set-top boxes as well as automotive. The systems will provide an intuitive user interface for access to information services such as stock quotes, news, weather, traffic and email -- all through natural voice commands.

In addition, Visteon is licensing SpeechWorks' text-to-speech (TTS) technology, known as ETI-Eloquence, to bring more personality and functionality to speech-enabled car services. With ETI-Eloquence, for example, car owners may hear, "Bear left at Route 1" or "You have one new email from John Smith." Dynamic information read to occupants in an intelligible voice is scheduled to include emails, directions, news, weather and traffic reports. Visteon plans to integrate ETI-Eloquence into its future telematics and multimedia products and services.

### **3.4.9.2 Webraska**

In 1997, Webraska patented a revolutionary technology for turning mobile phones into personal co-pilots. This technology, known as IbDN (Internet-based Distributed Navigation) combines high performance server-based algorithms and architecture with the most recent wireless internet device and cellular location technologies to offer the most comprehensive navigation solutions for all our partners. Webraska's Automotive Navigation solutions enables telematics service providers, car manufacturers and their partners to provide the car driver with an always up-to-date real-time navigation service optimized for the car environment.

Webraska has tested real-time traffic information, maps and navigation services for a number of firms including: PSA Peugeot Citroen, Vivendi, SAGEM. The concierge services will provide a virtual co-pilot to the driver. It offers maps along with a choice of itineraries with options such as quickest or least expensive. It provides journey-time estimations, even taking into account traffic conditions.

With their sights set on the telematics market, Webraska and ACUNIA have combined their areas of expertise, wireless navigation services and telematics technology, respectively, to offer car manufacturers wireless navigational services. Webraska has ported its IbDN (Internet based Distributed Navigation) technology in the ACUNIA telematics environment. Webraska uses the ACUNIA Service Development Kit (SDK) to develop and implement a distributed navigation service for use on ACUNIA's Open Telematics Framework, a Java-based open software designed for the entire telematics pipeline. This ensures that each hardware platform compatible with the ACUNIA technology will automatically be able to host the Webraska navigation services.

Webraska's IbDN technology allows car manufacturers to benefit from the data on its Global Navigation Server, which can be linked to any application, Internet provider, or external database. For the driver this means taking advantage of a dynamic navigation service that includes voice turn-by-turn directions that accounts for real-time

events, such as traffic and road conditions, and continuity of service to mobile phones, PDAs, and wired phones.

## **CHAPTER IV**

### **Market Dynamics And Forecast**

#### **4.0 Perspectives**

Speech technology continues to improve and is doing so rapidly. The improvements include better modeling techniques, increased robustness of the core speech recognition engines especially in accommodating to noisy environments, and larger vocabularies (stored utterances of words, phrases and accents). A number of the key speech technology firms have expanded the set of foreign languages that they can handle, which has increased to over two dozen.

Improved functionality in advanced speech recognition (ASR) technology is helping to drive a host of voice-enabled self-service applications. Current drivers that point to increased adoption include:

- Microprocessor technology is advancing to the point where speech recognition products can be very effective and accurate;
- Development tools such as VoiceXML make it easier to create new products; and
- Although systems are somewhat expensive, once the technology is installed, the cost of maintaining it is negligible. Additionally, prices are beginning to fall across all of the technologies.

In addition to these technology drivers, there are a host of end user issues. The major one is the increase in the population of wireless users. The mobile population consists of mobile phone users, mobile employees, CRM staff, and location-based services users.

## **4.1 Market Drivers**

The following section will provide insights into the market drivers, barriers and forecast for global telephony speech. The market drivers for this new area of speech technology and telephony include:

- Rising accuracy and falling prices;
- Wider adoption of wireless;
- New standards; and
- Regulatory issues.

### **4.1.1 Rising Accuracy**

Accuracy has been the specific focus of speech technology firms since it is so essential to their success. Speech technology has improved to such a degree that in many applications, the core speech engine can handle more than 100,000 calls per day with an accuracy level above 95 percent. This means that the engines are effective in most business and consumer settings today. There are several advances in speech science that are being utilized that continue to improve accuracy. These factors are:

- Self-Improving Technology – that is able to automatically adjust recognition mechanisms so that the speech engine improves its own capabilities to recognize words. This means, for example, that if a voice portal is serving an ethnic community, the software tools include artificial intelligence that lets it learn from all of the interactions and corrections it receives each day from callers. In this way, the speech engine learns how to recognize particular

accents and special regional pronunciations so that when later callers use the system, accuracy is improved.

- Location-Detection –the speech engine captures and uses the ANI information on incoming calls to detect the geographic location or even user group of the caller. This means that calls from 212 (New York) will pull up a different set of accent vocabularies than those from 214 (Texas). If a known subscriber calls into a voice portal, the subset of services they use is pulled into the database rather than the entire set, so that accuracy is higher.
- Larger Vocabularies – by building larger word and phrase vocabularies, speech engines can respond to a wider range of commands and can recognize application-specific words, such as stock names or ticker symbols for brokerage applications.

#### **4.1.2 Falling Prices**

Declining communication costs are fueling greater adoption of mobile services at lower prices. Lower prices are possible given the lower costs realized by carriers from efficient packet-switching architectures and enhanced network capacity. There have been reductions also in handset pricing, driven by falling component and modem costs. Expanded wireless network coverage and improved speeds and quality are due to the increasing deployment of digital versus analog networks. This means that wider adoption and expanded usage is happening in the very audience that needs speech technology the most – the mobile user.

The PELORUS Group asked the speech technology vendors what the two key drivers for their market were. All of them pointed to the trends of increased accuracy and falling prices for the technology. One key element in the lower costs is cheaper computer chips. Speech technology demands a lot of processing power so that today's hardware with higher speeds and large memories provide lower cost platforms for speech technology. This has resulted in a drop in the cost of speech systems from

nearly \$1 million in the early days to a range of \$100,000 to \$200,000 today. This driver will result in greatly increased adoption. Especially since speech server prices compare favorably with the cost of traditional deployments such as IVR.

As one example, Ascent provides a view of the cost of its AscenTel Server platform, which is used for voice portal applications, compared with a traditional IVR server platform. For an average size system with 20 menus and integrated database retrieval, the total cost of an IVR system would be in the \$270,000 to \$440,000 range compared with a speech-enabled platform between \$80,000-\$160,000.

<b>48-Port Server</b>	<b>Traditional IVR</b>	<b>Speech-Enabled</b>
Hardware/software	\$150,000-\$200,000	\$50,000-\$100,000
Development effort	\$120,000-\$240,000	\$30,000-\$60,000
Total for Deployment	\$270,000-\$440,000	\$80,000-\$160,000
Maintenance cost	\$60,000	\$20,000
Language used	Specialized tools	VoiceXML standard

Even the most expensive of the speech technologies, speaker verification, has dropped in price. Until recently biometric machines have been relatively expensive costing thousands of dollars per unit. In addition they have lacked the required speed and accuracy except for applications that involved special circumstances or requirements for extensive user training. However, more recently the situation has improved with the introduction of machines with introductory prices of \$1,500 per unit. It is expected that these positive trends will accelerate. Falling prices should allow more biometric technologies to be employed into medium security environments.

#### **4.1.3 Wider Adoption of Wireless**



Mobile populations are continuing to increase. They consist of the traditional mobile phone user, plus new populations related to mobile employees, CRM and location-based services.

#### **4.1.3.1 Mobile Users**

While nearly 40 percent of US households have Internet access, 96 percent have telephones. And 40 percent of the US population now has mobile phone accounts. Our forecasts show that this level will rise above 50 percent during 2002. Overseas the penetration levels for wireless services are amazing. In Iceland mobile phone penetration is 75.8 percent of the population and Finland's is 73.7 percent.

A number of market studies show the need for greater convenience for mobile phone users. These studies show that mobile subscribers want voice dialing. And consumer demand for safety and regulatory action are driving the adoption of hands-free sets in automobiles. Consumers who are on the go demand easier to use interfaces. As they start using more data services in addition to voice, the mobile phone itself becomes a barrier. Nuance Communications, a speech technology firm, released a study that shows that 80 percent of wireless users prefer voice recognition to touchtone entry. Among users of speech recognition systems, 95 percent of them were satisfied with the experience.

Another important market segment are mobile workers. Enterprise solutions enable employees, customers and company partners to use mobile handheld phones or devices to connect with each other, corporate databases, applications and intranets. For example, United Parcel Service has implemented a wireless shipping information application. The network can format the UPS customer data for almost any device, including two-way pagers, Palm PDAs and Web-enabled phones. This enables delivery persons to receive up-to-the-minute alerts on deliveries and changes on any customer in real-time.

#### **4.1.3.2 Mobile CRM**

Customer relationship management (CRM) applications for the modern contact center are a key driver in that market. Included in this vision is the provision of mobile CRM or wireless CRM. Mobile CRM provides access to the enterprise's knowledge base and makes use of advanced mobile hardware and software to deliver the information. Mobile CRM is aimed at improving the relationship with customers and with employees.

Employees can dial into or call into the enterprise database to access the records they need for either sales force automation applications or field force automation (technical service). This includes the addition of a voice-activated interface, along with WAP data entry and portable laptops. Useful applications are being developed by companies such as PeopleSoft, which has developed Mobile Sales for WAP software, which gives field sales representatives access to enterprise information over WAP phones. This application lets sales representatives view and update information on companies, contacts, tasks, leads and opportunities using Openwave System Inc.'s microbrowser. Mobile Sales for WAP follows the PeopleSoft's release last year of Mobile Field Service for RIM, which allows dispatch centers and field service representatives to exchange information through messages on Research in Motion's interactive pagers.

Speech recognition is also gaining a hold in call centers as a self-service tool that the customer can use to either find a live agent or to extract information from a database. These applications make it possible to introduce a much greater range of functionality than was available through traditional touch-tone IVR. Voice applications allow callers to quickly progress through options using natural language rather than tediously progressing through long menu selections. However, ASR applications that merely substitute a spoken number for a touch-tone key press do not optimally exploit the technology. To get the most benefit from ASR systems, an entirely different menu flow and approach must be employed. ASR applications require specialized design and fine-tuning to be successful.

### **4.1.3.2 Location-Based Services**

The near-term introduction of location-based services will be a major driver for adoption of speech in mobile environments. There are two key but very different drivers related to adoption of location-based services. One driver is the regulation requiring carriers to implement Enhanced 911 and the other driver the development of innovative new services for commercial, non-emergency location-based services. The dynamics of these primary drivers will result in higher deployment levels of location-finding technologies and greater potential for location-based service revenues.

From the year 2002 and forward, there will be large-scale implementation of location-finding technologies into the US wireless networks due to adoption of Enhanced 911 (E911). Other countries have also established improved wireless 911 capabilities as a priority. With an improved, more accurate, location-finding infrastructure, service providers will be able to move ahead with development of much more accurate emergency response services as well as commercial location-based services and applications.

There is still much speculation about what revenue opportunities will result from providing location-based information and services. There is no lack, however, of discussion about the myriad of marketing opportunities based on or enhanced with location-based data. The primary applications would include information services, telematics, fleet and vehicle tracking, people tracking, and asset tracking. Additionally, location data could improve Internet portals in the area of information services, push and pull advertising, directions and navigation services, such as traffic status. Another potential area relates to operator services: enhanced directory assistance, including yellow pages, concierge services, events notification, and points of interest and local attractions services.

### **4.1.4 Speech Technology Standards**

As technologies such as automatic speech recognition and text-to-speech conversion are gaining acceptance, VoiceXML has emerged as the standard for voice-activated applications. Voice Extensible Markup Language (VoiceXML) is a standard language for building interfaces between voice-recognition software and Web content. Just as Hypertext Markup Language (HTML) defines the display and delivery of text and images on the Internet, VoiceXML can translate any XML-tagged Web content into a format that speech recognition software can deliver by phone.

VoiceXML enables any computer-stored and text-based content to be made available through telephones and other devices using both speech and ordinary touch-tone interfaces. It effectively functions as a touchtone replacement that lets the customer speak to the system to retrieve information, order products or brochures, check the status of orders, and take customer surveys. HTML helped create a visual channel for conducting transactions on the Internet through a Web browser. VoiceXML helps deliver an audio or voice channel for doing the same thing.

#### **4.1.4.1 VoiceXML**

One of the most important developments aiding in the adoption of speech technologies is the success of the VoiceXML standard. The VoiceXML Forum ([www.voicexml.org](http://www.voicexml.org)) boasts some 600 members. There are several thousand application developers using VoiceXML today. For example, VoiceGenie's Developer WorkShop now hosts some 3,000 applications. Their development platform provides a foundation for creating and testing voice applications using a choice of ASR and TTS engines.

In order to achieve wider acceptance of the VoiceXML language, in April 2000 the Forum submitted it to the World-Wide Web Consortium (W3C) for consideration. The W3C's Voice Browser Working Group eagerly took on the job of review and revision. The result has been VoiceXML 2.0, whose first Working Draft came out in January 2001. There are two more Working Drafts planned, and several more steps until VoiceXML 2.0

becomes a full W3C Recommendation. The Working Group hopes to have the process completed in early 2002.

Used in telephony speech applications, the VoiceXML standard for speech recognition has accelerated the development process. Using VoiceXML, developers can create VoiceXML-enabled Web pages that anyone in the world can download and listen to with VoiceXML browsers. These VoiceXML browsers or interpreter operate just like Microsoft Internet Explorer or Netscape Communicator but focus on speech audio not visual. An increasing number of firms are now selling VoiceXML interpreters. These include the major speech software firms profiled in this report, such as AT&T, IBM, Nuance, SpeechWorks, Motorola and Unisys.

Voice portal firms have adopted VoiceXML browsers into their systems. Acting as ASPs for portal creation, firms such as BeVocal, General Magic, HeyAnita, Tellme and others, are providing developers with access to their systems to work on creating application programs in VoiceXML. As interpreters are made more widely available, development of speech technology applications will increase.

#### **4.1.4.2 Other Standards**

A variety of other standards efforts are underway in the area of speech technology. They are:

- Speech Recognition Grammars – various efforts are underway to standardize grammars and language syntax use among speech vocabularies;
- Natural Languages Semantics markup Language – similar to VoiceXML in focusing on natural language meanings related to individual natural language utterances;
- Speech Synthesis Markup Language – standard that would help in synthetic speech (text-to-speech and speech-to-text) applications in the Web; and
- QSIG – a standard that enables interoperability among PBXs and other voice equipment related to signaling and feature interoperability.

#### **4.1.5 Regulatory**

One of the major drivers for adopting speech technology in wireless markets is regulatory pressure. There are concerns that increased mobile phone use is causing more automobile accidents. The reasoning behind legislative efforts is that handheld mobile phones are distracting to the driver and can cause accidents. To support their cause, advocates for controls cite recent studies that indicate cell phones may provoke driver inattention and cause accidents. The most cited report is one from the New England Journal of Medicine (1997), which reports that drivers are four times more likely to have a crash while using a cell phone than when they are not. It is also reported that more accidents are caused by drivers on the phone than by drunk drivers.

The move toward stricter control of mobile phone use in vehicles is international. New legislation in countries including Korea, Australia, and Israel ban the use of handheld phones in moving vehicles. Thirteen countries have similar bans, including Spain and Italy. And in the US, various municipalities, counties, and states are considering action. Already New York State's new law has gone into effect. It bans the use of cell phones while the car is in motion.

Since 1995, 37 states have proposed bills regulating cellular phone use in automobiles, but only New York State has passed one. Given the success in New York State, some twenty other states considered action again in 2001. According to the National Conference of State Legislatures, recent reviews have taken place in New Jersey, Pennsylvania and Delaware. Broader bills to study driver distractions have been enacted in four states including Pennsylvania, which recently completed its study.

Although the carriers have fought legislative action in the past, this has changed. Verizon came out in support of statewide laws to require motorists to use hands-free sets. This of course would mean new equipment sales for wireless firms, selling phone kits that include a microphone, a speaker and a car mount for the phone.

On the Federal level, a ban on handheld cell phones was introduced by Senator Corzine of NJ and Representative Ackerman of NY. It is still pending in Congress. The US National Highway Traffic Safety Administration is continuing its investigation into whether something should be done to minimize the distraction caused by cellular telephones and computerized gear in cars and trucks. It has determined that at any given time during daylight hours, three percent of all drivers of cars, vans, and pickups are on cell phones.

Agency officials, though, do not believe they have the legal authority to issue a federal rule restricting drivers' phone use. However, in a recent survey (November, 2000) conducted by NHTSA, it concluded that "voice recognition is a viable alternative" to reducing potential hazards that in-car phones and telematics systems might pose. This is a strong impetus for adoption of speech technology for mobile applications.

## **4.2 Market Barriers**

Decisions about strategy are critical for any carrier considering implementing speech recognition. First, carriers must determine whether the services will be offered as a tool for increased demand and loyalty or as a means for revenues itself. The carriers must also work with the technology to ensure its acceptance among consumers. The carriers must keep the menu options simple enough so users find the service efficient and useful. As with IVR, if too many choices are given to users, they will be confused and have difficulty. If they have difficulty the first time they use it, they may not come back.

Another possible problem is the quality of service. If the voice portal cannot handle the voice accurately or if noise in the background disturbs the call, consumers will be frustrated. If the voice portal does not accept the voice command the first or second time, the customer will quit. This is especially true for services that will charge subscription fees.

These concerns and others are covered in the section below. It contains information on quality issues, usability issues, assessment of real potential, and mobile Internet.

### **4.2.1 Quality Issues**

Applications that merely substitute a spoken number for a touch-tone key press do not optimally exploit the power of speech technology. To get the most benefit from speech recognition systems, an entirely different menu flow and approach must be employed. ASR applications require specialized design and fine-tuning to be successful. The most common concern related to speech recognition is accuracy. Levels of accuracy continue to rise and have been variously reported from 95 percent to 99 percent. The main errors that can occur are:

- Rejection errors – when the speech software cannot recognize the word as spoken by the user; the user says, “eighty” and the system responds, “Sorry, I cannot understand you, would you say that again;”
- Substitution errors – when the system incorrectly selects a different word than what was spoken by the user; the user may say “send” and the system hears “end;” and
- Spurious response errors – when the system hears extraneous noise or a cough and interprets it as a word; a user sneezes and the system hears the word “stop”.

Noisy environments, a user’s cold, or accents may disrupt the proper working of the system. Since performance will vary depending on the demands of the application, there should always be an opt-out process. This would allow the default use of touchtone or opt outs back to the operator in cases of multiple failures.

### **4.2.2 Usability Issues**

In replacing IVR applications, the goal of speech application is to move toward natural language. However, in most cases today, users still have to talk to speech recognition systems in a relatively structured way. The old systems had low levels of



interaction characterized by the way one would talk to a small children, simple-minded folk, or foreigners who do not know our language too well. This meant speaking more slowly and repeating words if the spoken information command was not understood the first time. Newer systems may not be natural language, but they present a much more natural interaction.

Depending on the sophistication of the speech deployment, speech may not necessarily be the fastest way to interact with unified messaging functions, compared to a touch-tone key, a screen-press, or even a keyboard input command. However, there are times when one needs to be "hands-free" to control communications.

The other side of speech technologies is in the system reply – the spoken replies given to the user. These text-to-speech applications have also made progress, with the traditional "drunken Scandinavian" voice giving way to a more natural-sounding speech. Most voice-response applications still rely heavily on pre-recorded information, rather than TTS, except where the information is so variable that it cannot be pre-recorded, e.g., names. Most people find it difficult to listen to synthesized speech for very long because it is an effort and a strain to follow. However, as long as the speech is easily understandable and limited, selective use of TTS works.

### **4.2.3 Assessment of Real Potential**

Right now there is a lot of hype about the various mobile markets and their potential. This includes the areas of speech technology. Having over promised in the past, technology vendors are now more cautious in what they promise. But carriers have to be realistic. Right now most of the major technical barriers have been removed. Speech technology has finally come of age and permits usable voice-activated services. Quality is still in need of improvement, but has grown tremendously. However, the set of services offered are still small in number and have not been fully assessed as far as market demand.

The voice portals and wireless Web sites do offer a variety of services, but nothing that can approach what the real Web has to offer. In time, more will be added, but realistically only a small subset of Internet information can be applied to the mobile situation where the amount of time people have to listen to information read to them is limited. And there are no graphics, of course. So most content applications will remain focused on quick facts, messaging and the ordering of tickets or services that someone does not have to see, such as airline tickets. Initial success overseas with Omnitel and NTT DoCoMo and here in North America with Sprint PCS and Bell Mobility already suggest that the successful categories are chat, email, games, entertainment and e-commerce.

#### **4.2.4 Mobile Internet**

Mobile Internet is also be widely touted as the next big thing. But just because 3G networks and Web-enable phones are possible technically does not mean that users will demand it. The market in Japan is hot right now, but the usage elsewhere has yet to be determined. Caution might be advisable for several reasons. One is that the types of handsets that are in demand are smaller and lighter. But if Web is to be used, there must be larger and have better screens. Which way will it go? The compromise may make using the Web a difficult task. Unless of course all interaction is by voice command!

The second reason for caution is pricing. Information is largely free on the Internet, but using the Web on a cell phone costs money because of airtime. And the third reason for caution is the increasing availability of the Web. The penetration of household and office PCs and Internet connections has reached the mass market. The vast set of information is already available at an easy mouse click away. Can the reduced content of Web phones really satisfy the users? Probably not, but unique information and e-commerce solutions may.

### **4.3 Market Size and Forecast**

The market for speech technologies is in its beginning stages. Greater potential is indicated by the success of the initial set of players (profiled in Chapter III). In just the last two years, the robust improvements in the core speech engines have made them vastly more useful. Plus, the successful application of speech to some key industries, such as financial services, bodes well for this marketplace.

### **4.3.1 Revenues From Key Players**

The major revenue source for this marketplace today is the revenue coming from the major speech software firms. These players are the ones that sell the core speech engines that translate the incoming spoken voice to something that is useful to a particular platform (e.g. telephony) or application (e.g. voice portal). Sometimes they are referred to as speech engines, which means they are the locomotion behind the accurate translation of spoken sounds into recognizable words. Chapter I gives details on how this works. There are four types of speech engines: automatic speech recognition, text-to-speech, speaker verification and natural language processing. The major portion of the speech software firm revenues comes from licensing their software to partners, system integrators or even end customers to build various speech-enabled solutions. We refer to the revenues coming from the speech engine firms as the speech software revenues.

Figure 4.1 shows a chart of the top twenty firms globally and how they are classified by type of firm. Are they a public or privately held company? Is their company focused on speech software exclusively or are they a division of a larger enterprise. Are they US or foreign based? The PELORUS Group identified twenty firms as the major players in this arena. The figure shows thirteen US-based companies and seven foreign-based. It includes five private firms and fifteen public companies. There are nine firms that are largely speech focused and eleven that have speech software as just one of several product lines.

As stated at the outset, telephony markets are the dominant focus of the core speech technology firms today. There are numerous announcements every week about new telephony service providers or enterprise call centers launching a new speech application. For example, in 2001 alone, Nuance had a number of major wins. Table 4.1 shows the deployments in the telco carrier segment for Nuance in 2001, which includes major wireline and wireless carriers in the US and Canada.

**Table 4.1**  
**Nuance ASR Deployments in the Telco Carrier Segment**  
**As of January 28, 2002**

<b>Carrier</b>	<b>Application / Service</b>	<b>Partner</b>
AT&T Business	Enterprise customers	Tellme
AT&T Business	Toll-Free DA	Tellme
AT&T Wireless	Voice portal	Tellme
Bell Canada	Voice portal Bell Mobility's VoiceNet	BeVocal
BellSouth	Voice portal	BeVocal
Cingular	Voice portal	BeVocal
Qwest Wireless	Voice activated dialing	BeVocal
Qwest Wireless	Voice portal	BeVocal
Sprint PCS	Call center customer care Uses Nuance's "Say Anything" technology	Edify (IVR) and Viacore (Apps, system integration)
Sprint PCS	Voice activated dialing "PCS VoiceCommand"	
Telus	Enterprise services	

#### **4.3.2 Revenue Streams**

The PELORUS Group's forecast for global speech markets was developed in three steps. First, the revenues from the core speech software vendors were calculated for two

base years, 2000 and 2001. Figure 4.2 shows the revenues for these two years, which equaled \$398 million in 2000 and \$483 million in 2001.

Revenue from this set of players is fairly straightforward because it comes from the fees for licensing the cores speech engines (software) and professional services for assisting in its deployment. In recent SpeechWorks public announcements, for example, it simply divided its revenue streams into three components:

License fees	60%
Professional services	37%
Resale	<u>3%</u>
	100%

Seven of the key firms in this group are focused on selling speech software only -- Fonix, Nuance, SpeechWorks in the US and Elan, LocusDialog, SRC and Telisma outside the US. For the others in our set of core firms, the revenues are much more difficult to count, because their speech units are buried within larger companies as divisions, units, or subsidiaries. If revenues were not reported, then revenues were calculated using industry benchmarks such as revenue per employee, average size of contract and market share. This method was used to calculate revenues related to units such as Lucent Speech Solutions Group, Philips Speech Processing, or InfoSpace's speech component of its Wireless Division.

#### **4.3.2.1 Speech Software Revenues**

To calculate the entire revenue stream from the existing set of speech software vendors, one must factor in the other vendors not included in our group of core firms. This is done using the industry benchmark of 80/20. The top players in any market generate about 80 percent of the total revenues and the rest of the players generate the other 20 percent. Therefore, total revenues are calculated by dividing by .80 to get the total that would include both the core firms and all others. These figures are given in Figure 4.3. It shows that revenues from all speech software firms were \$497.5 million in

2000 and \$630.75 million in 2001. These revenue figures are used as the base years for all further calculations.

Figure 4.4 shows the PELORUS GROUP estimate for growth rates for speech software revenues forecasted to 2006. This is based on recent growth rates and on estimates given to us by the industry players. As one sees, there was a lower rate in 2001, given the slowing economy and the impact of the terrorist attacks on 9/11. However, it was still a substantial 21.4 percent. But outlooks are for improved adoption and revenues going forward. Growth rates will increase above the thirty percent (32.5 percent) level by the year 2004.

Using these growth rates, total speech software revenues can be calculated. Figure 4.5 shows the estimates for total revenues derived from firms selling speech software for the years 2001 to 2006. In year 2001, the revenues are \$604 million, reaching above \$1 billion by 2004 (\$1.279 billion) and reaching a high of \$2.5 billion at the end of our five year forecast period.

#### **4.3.2.2 Other Market Components**

The revenues from the speech software firms comprise the majority of revenues in the marketplace today. But by focusing only on the sales of speech software, one neglects the other platform, application, and service revenue streams that come from the application of speech technologies in the various markets. Since this is a fairly young market, these other market components are still only a small portion of the total at this stage. However, there are signs for a significant movement in both breadth and depth of application in a diversity of market segments. These must be factored into the mix for the revenue forecast.

As profiled in Chapter III, these other firms consist of application, platform and service firms who incorporate speech software into their application areas. Figure 4.6 shows a chart of ten applications areas and the firms that are involved in each of them. Some of the firms are active in several of the application areas. These firms have

specialized expertise in these application areas, such as directory assistance or enterprise auto attendant. They license the speech engines from the speech software vendors and partner with them to develop and sell solutions to their specific markets. The firms listed in this chart are only a few of the many firms operating in these market areas, but they were chosen since they have been involved at least a little in integrating speech software into their solutions.

So far, widespread adoption of speech technology in most of these applications areas has not taken place. The PELORUS Group has completed major industry studies in a number of these areas (see [www.pelorus-group.com](http://www.pelorus-group.com)). In reviewing the findings from these reports, one sees that only a small subset of firms have already updated their product lines with speech technology. Some of the key firms that have been profiled in Chapter III above. Some, though, have launched major new initiatives in speech technology. And all of the reports indicate that speech technology applications will be a critical trend for the future.

### **4.3.3 Total Market Revenues**

To calculate the impact of these other firms, one must apply some model that reveals proportion of revenues by type of player. To do this, one can take the value chain model and apply it to assessing total market revenues for the global speech market. Several models were given in Chapter III, which provided insight into the key stakeholders and their roles. For example, see the IBM value chain model in Figure 3.2. In discussing this model with the industry players, a consensus was reached. The most applicable model comes from the software industry, whose revenues benchmark was created from years of experience in observing software markets in action. At market maturity, software market revenues divide in this way:

FIRMS	SHARE
Service Providers	50%
Hardware/Platform/Apps	30-35%

Software	<u>15-20%</u>
	100%

At market maturity, therefore, one can assume that software would be 20% and the combination of hardware/apps and services would be 80%. As stated earlier, the global speech market is in its early stages so software would dominate. Today, one can assume that only 10 percent of the market is related other firms (hardware, platform, applications, services) and 90 percent would be for software. This will change over time and given the rapid development of speech applications, it will probably do so quickly. Figure 4.7 shows the evolution in revenues by source for the period 2000 to 2006. Starting with 90 percent of the base in 2000, software decreases in overall contributes to 65 percent by 2006 and all other players contribute 35 percent of the total.

Using these sets of numbers, one can take the revenues developed for speech software firms as the base and calculate total speech market revenues. The total speech market revenues would include revenues from speech software firms, plus all others that provide applications, services and specialized platforms based on speech software. These numbers were developed for Figure 4.8, which shows revenues of \$693.97 million in 2001. The market will move above \$1 billion in 2003 (\$1.222 billion) and then grow to \$3.869 billion at the peak of our five-year estimate in 2006.

The growth rates associated with the total speech market are shown in Figure 4.9. The lower rate for 2001 is explained by the economic depression and effects of 9/11. Rates will continue to move back up as the economy rebounds, moving into the 30 percent range for years 2002 through 2004 and rising above 50 percent for 2006.

The hardware/services component of this market is shown in Figure 4.10. This shows revenue growth from only \$90 million in 2001 to \$1.4 billion in 2006.

#### **4.4 Revenue Models**

All of the key stakeholders in this market decry the lack of clear revenue models.



The lessons from services deployed in the US and Europe are just starting to emerge, however. But much more history and market experimentation must happen before a clear path emerges.

#### **4.4.1 Advertising**

History tells us that advertising is a driving force behind every established media including radio, TV, newspapers and the Internet. This was thought to be the case as well for wireless data services. Advertising on WAP and speech-enabled voice portals was launched in the early deployment phases of these services. But that has changed and they are largely missing from the current revenue mix for location-based services. However, with careful management new forms of advertising might work.

To be successful new forms of mobile services, such as location-based and telematics must be an opt-in service, letting consumers choose what data, alerts or ads they will receive. The danger is that despite this assumption, aggressive marketers will blanket users with unsolicited messages anyway. Using location-finding techniques, LBS means that businesses will be able to send promotions to customers when in specific geographic zones. Real-time personalized marketing is possible with LBS. The government has weighed into this debate as well. Data collected from cell phones is protected by FCC rules, which require all carriers to treat that information, including location, as confidential, unless customers give express permission to share it with third parties, or in emergencies. But these rules do not apply to data collected by telematics services, such as OnStar.

It can be said that consumers perceive alerts that are relevant, compelling and convenient as content rather than advertising. In addition, opt-in permission based models can be effective models for advertisers. Most consumer studies reveal that there is a fine line between providing consumers with a positive experience and turning them off with irrelevant messages or intrusive wireless spam. Consumers are annoyed by ads that simply tout a product, with no interactive options or corresponding offer.

Additionally, participants are irritated by alerts that are not targeted to their specified interest categories. These findings illustrate that wireless marketing is a double-edged sword; when done correctly, it is highly effective, but when abused or shoddily planned, it can truly alienate and incense consumers.

For this reason, if wireless marketing models are executed irresponsibly, the adoption of wireless marketing will be in peril. Carriers, the ultimate gatekeepers, will fear customer dissatisfaction and abandon any initiatives they believe might cause churn. In addition, it is safe to assume that government agencies such as the FTC and FCC will be quick to offer regulations to protect the rights of consumers from rogue marketers that violate privacy rights.

Therefore, one concludes that if it is done responsibly, adhering to a permission-based model, wireless advertising has great upside potential for carriers, publishers, advertisers and consumers. For carriers and publishers, it can be an effective way to increase incremental revenue, enhance customer loyalty and reduce customer churn. For advertisers, it represents an effective means to drive consumer response and enhance brand awareness. And for consumers, wireless advertising is the ultimate one-to-one marketing tool, providing a value added service that offers convenience in gaining access to products and services that they want.

#### **4.4.2 Subscription**

Qwest Wireless offers its Voice Browsing service, which is priced at \$5 per month. When testing the service, Qwest found that most customers were willing to pay a nominal fee with the guarantee that they would not be subject to advertising. Currently most independent portals, like Tellme, are advertising supported. All the details of the financial arrangements are not public, but it is known that Qwest and BeVocal share the subscription fees. Qwest is BeVocal's first carrier partner.

The price model of about \$5 per month is also confirmed by the AOL by Phone service, which American Online recently began offering. It is priced at \$4.95 per month.

Yahoo! offers its basic service as free, but charges for premium services. And it bills for extras such as disk space or when users patch through its service to make long distance calls.

The model of subscription fees rather than advertising-based solutions is also confirmed by a study published in Telephony magazine that notes consumers expressing reduced interest in voice portals with ads. The carriers offer wireless Web as one of their data services.

#### **4.4.3 M-Commerce**

In the long term, m-commerce will be a major driver for mobile services and applications. Some of them will depend in large part on location-based mechanisms. Given the infancy of m-commerce today, the potential in this area is hard to predict. Much remains to be done in the area of developing secure, easy-to-use financial transaction mechanisms.

For purposes of assessing the potential of mobile commerce, it is fair to say that totally advertiser-sponsored services are unlikely to play a large role. It is even more important to exercise caution when advertising to business people. It is thought that a business user might tolerate a single coupon or a one line special notice from their favorite store, if they obtain data, such as the nearest store location sent to a mobile device for free. But most professionals will bridle about advertising that fills their mobile device screens. Even a few unsolicited and unwelcome generic advertisements to a business user that comes without explicit consent could seriously tarnish a hard won customer relationship.

#### **4.4.4 Effect Cost Reductions**

The focus of this report is on market opportunities, but we cannot ignore the opportunity to reduce costs. Customers can achieve significant cost reductions by

adopting speech technology into their call centers, for example. With user-friendly voice interfaces, customers can navigate to the information or action they need to take. Reductions in time spent on the call and, in some cases even eliminating live operators, wireless carriers can reduce costs.

Service bureaus that operate some of the audiotext services may realize cost savings by becoming voice-activated. Using speech-based data storage, they can save transmission costs by pulling in VoiceXML pages over IP networks instead of transferring the call to public network-based services. Some of the core IVR applications may migrate to the more dynamic voice portal model. It is fairly straightforward to convert traffic that was going to one 800 number so that it travels through a voice portal server instead. Callers will not know the difference, but their calls may be handled over cheaper IP networks. In the future, call center calls will be served by voice ASPs that run VoiceXML servers that pull up VoiceXML pages of data from across cheaper IP networks.

NetByTel, a Florida-based wireless ASP, has conducted ROI studies with some of its customers documenting cost savings. It has helped its customers to implement speech technology in their call centers. NetByTel has seen its customers realize reductions in the cost of taking phone sales orders and customer service calls. NetByTel's main product is NetByTel Connected, which is designed for call centers and is based on SpeechWorks technology. Using the system, callers making calls into customer service centers can navigate to the appropriate information or execute transactions using voice commands. NetByTel's customers include Priceline.com, Office Depot, and BigStar.com.

Office Depot reports that handling customer calls through a voice interface costs about 84 percent less than routing calls to live customer service operators. Office Depot customers can call 1-888-GO-DEPOT and through a series of voice prompts can order a catalog, locate a store in their area or even buy products through an automated voice-enabled system. Office Depot said that they had to spend only about 16 hours of technical effort to deploy the NetByTel system and it was tested and up and running in less than 30 days.

Plus by upgrading their operations with speech recognition, call centers can improve customer satisfaction. Call center interactions for the customer may be frustrating in IVR set-ups. Callers may have to plod through multiple menus of touchtone and IVR menus. Sometimes after they reach the right spot, customers are still put on hold. Or they zero-out to a live attendant. To handle overload situations, call centers either hire more staff or outsource the overload. This costs between \$1 and \$15 per call. If new speech recognition technologies are used the costs can be reduced down to \$.10 to \$1 per call. Thus cost savings are achieved anywhere from \$.90 to \$14 per call.

The automation of telematics services is also a source of cost reductions. Voice activation will save money for telematics firms and automakers. Today telematics service providers use customer service representatives to answer calls from subscribers. ATX is partnering with IBM to test technology to relay the driver's verbal request to an ATX server, where they would obtain information directly from the database using voice commands. This allows ATX to handle more calls without hiring more operators. By storing information at the call center, this reduces the amount of electronics that need to be installed in cars, so automotive manufacturers save.

## **4.5 Market Shares**

Since this market is large and complex, it helps to analyze its key components to gain a better perspective on the market dynamics. The sections below provide insights on the share of the market by customer markets, geographic focus, and technology.

### **4.5.1 Customer Markets**

The focus of this report has been on telephony applications in the speech technology arena. But which of the customer markets will dominate in the future? This question was asked of the speech technology vendors in this study. Figure 4.11 shows the view for today and 2006. Most vendors think that enterprise deployments dominate today's

market with 40 percent of the total, 35 percent for carrier and 25 percent for embedded (device and automotive). However, by 2005, carrier deployments will lead with 45 percent of the market, 35 percent for enterprise and 20 percent for embedded applications such as mobile devices, phones and telematics.

One can then apply these ratios to the revenue estimates. Figure 4.12 uses the ratios for the years 2001 and 2006 and applied them to speech software revenues shown in Figure 4.5. The resulting estimates in this figure shows that carrier revenues rise from \$151 million in 2001 to \$1.13 billion in 2006, becoming the largest segment in that year. Enterprise revenues are estimated to be \$241.5 million in 2001 and \$880.28 million in 2006. The smallest segment is embedded markets, which contribute \$150.94 million in 2001 and \$503.02 million in 2006.

#### **4.5.1.1 Killer Application**

It is important to realize that this market is characterized by a diversity of applications, with no one "killer" application dominating the market. In the enterprise customer market, the applications are straightforward, since the corporation is automating the straightforward applications of IVR updates, the switching/routing of incoming calls and voice mail enhancements. On the carrier side though, demand from end users will determine which applications are pursued. There are so many different consumer applications that it would be most unlikely that one would emerge as the sole winner. The Voice XML Forum reports on the top speech applications in use today. Figure 4.13 shows figures for the most used consumer portal applications, revealing that voice-activated weather reports is the most often cited application. Speech-enabled voice mail, email and dialing (VAD) also rank in the top five. Financial news and sports news are the other two applications in the top five. Although none of these applications may be pervasive, clearly consumers find many different uses for speech technology.

#### **4.5.2 Geographic**

Speech technology has been deployed the most so far in North America. Figure 4.14 shows that vendors estimate 65 percent of their contracts are in North America, 25 percent in Europe and 10 percent in the rest of the world. This will change over the forecast period, with Europe increasing to 35 percent and the rest of the world, led by Asia-Pacific, increasing to 20 percent.

One can then apply these ratios to the revenue estimates as was done above for the speech software revenues. Figure 4.15 used the ratios for the years 2001 and 2006 and applied them to speech software revenues shown in Figure 4.5. The resulting estimates in this figure shows that the dominant market of North America generated \$392 million in 2001 and is expected to generate \$1.131 billion in 2006. Europe generated \$150.94 million in 2001 and will contribute \$880.28 million by 2006.

### **4.5.3 Technologies**

When questioned about which of the speech technologies sells the most, the vendors we interviewed said that automatic speech recognition or ASR was dominant. Figure 4.16 shows that the consensus view to our question. Overall, it was found that ASR accounted for 60 percent, TTS for 25 percent, speaker verification for 10 percent and natural language understanding for 5 percent. These ratios will probably remain the same during the forecast period.

One can then apply these ratios to the revenue estimates. Figure 4.17 used the ratios for the years 2001 and 2006 and applied them to speech software revenues shown in Figure 4.6. The resulting estimates in this figure shows that the largest segment, ASR revenues, rose from \$362.25 million in 2001 to \$1.5 billion in 2006. TTS revenues were estimated to be \$150.94 million in 2001 and will rise to \$628.77 million in 2006. Speaker Verification (\$251.51 million) and Natural Language Understanding (\$125.75 million) are the smaller segments in 2006.

## **CHAPTER V**

### **Future Industry Trends And Issues**

#### **5.0 Perspectives**

This is a young and dynamic industry. Applications based on speech technology are finding their way into a number of markets and industries. This report concludes with a discussion of the future industry issues and next generation technologies that will impact this arena in the coming five years.

#### **5.1 Industry Issues**

The speech technology market is fast moving, with many new initiatives and directions. Standards are an issue, since wider application of speech technology depends on a widely available applications development platform. Many of the issues for speech, of course, relate to the wireless markets – especially wireless data services and telematics. The key developments emerging in the speech markets are:

- The new standard, SALT;
- Enhanced DA Services;
- Wireless Data Devices;
- Voice Commerce;
- Telematics; and
- Voice over IP (VoIP).



### 5.1.1 SALT

VoiceXML's role as the pre-eminent standard in speech technology is being challenged. A coalition of leading technology companies has introduced a new standard for accessing Web services through voice commands. Spearheaded by Microsoft, industry players unveiled a proposed standard for developing voice-activated Web applications, called Speech Application Language Tags (SALT). In addition to Microsoft, founding members of the newly formed SALT Forum include Intel, Cisco, Philips, Comverse and SpeechWorks International.

The SALT standard is based on HTML (HyperText Markup Language), XML (Extensible Markup Language) and other Web development standards. It will be one of many technologies designed to facilitate a widely touted concept of delivering information to users at any time and on any computing device. The initial version of the specification will be available in the first quarter of 2002 and will be submitted to a standards body for review by the middle of the year.

In addition to creating a standard method for sending and receiving voice messages, SALT will allow hardware and software makers to create a standard user interface for hosting new voice services on computing devices. The SALT Forum will work with companies to build a speech interface onto a Web site so it can be accessed by telephone or mobile device. The technology is intended to make Web services accessible through a telephone or any computing device that can transmit voice. With the technology, these companies say voice will replace the functions of a keyboard, telephone keypad or handheld stylus.

Microsoft hopes that SALT will enable its tens of thousands of third-party application and Web site developers to add speech capabilities into their applications using their existing Web programming skills. The technology will be supported in Microsoft's Internet Explorer Web browser and will be part of Microsoft's emerging .Net initiative for delivering services across the Internet. Microsoft said it would incorporate

SALT into its Visual Studio.Net offering as the technology evolves so developers can build voice-activated applications within its development software suite.

However, SALT is designed to work independently of the operating system and on any computing device that embraces the standard. The technology will also be free for use by Web developers. Although the idea has gained some industry support, Nuance and others think that it may slow down the development of the currently widely used standard, VoiceXML (Voice Extensible Markup Language).

Those working on the SALT project, though, say it is expected to be a lightweight alternative to VoiceXML. They say VXML and SALT can coexist, since they are driven by two different models. SALT is oriented toward modifying existing HTML pages, cell phone and PDA screens, and multimodal devices. SALT's target application is Web page development that is enhanced using special tags for voice prompts and speech recognition. VoiceXML, on the other hand, is oriented to telephony systems, where high volume telecommunications applications are upgraded to handle voice-activation. VoiceXML is used to upgrade IVR systems and enable voice dialing services.

Using SALT, developers will only need to add simple, lightweight SALT tags to existing applications, rather than rewrite the application. It will simplify the creation of multimodal applications where users input and receive with both their voice and their keyboard. It is really designed as a tool that extends the standard Web page into the speech-enabled world. How do SALT and VXML compare?

- VoiceXML is a server-based and is a computer-telephony solution, but SALT is Web server/client-oriented.
- SALT is simpler than VoiceXML. SALT follows HTML programming models, so anybody who can setup Web pages should be able to SALT-enable them. VoiceXML is more powerful and complicated; one can build complete IVR systems based on VXML, but instructions must be structured correctly.
- VoiceXML is aimed at the computer-telephony channel and the VARs who have expertise in its complexities. SALT is aimed at Web developers.

Vendors that support both VXML and SALT, like Intel/Dialogic and Cisco, indicate that the two will co-exist in the marketplace, typically supporting different channel niches and may even co-exist within solutions.

### **5.1.2 Enhanced DA**

Another important area for speech technology is Enhanced Directory Assistance. Enhanced DA is viewed as a major new source of revenue for operator services units and DA providers (see The PELORUS Group's Enhanced DA report). The providers in this market are adding voice-activation for all of the enhanced DA services: National DA, Reverse Look-Up, Category Searches, and Concierge Services.

However, according to speech technology vendors, there are some barriers. Speech recognition in large directory assistance databases is difficult to implement. This is mainly due to the sheer volume of listings that a system must be able to recognize. Several companies are working to overcome the problems.

For example, Phonetic Systems of Burlington, Massachusetts has developed a phonetically-based system that offers a National or Global DA service called Directory-Assistant. It claims to work with databases of more than 500 million listings. However, Phonetic Systems and others do not yet have nationwide systems for DA up and running. Some of the vendors are operating trials.

Nuance claims that it is close to offering a partial solution. Since 80 percent of directory assistance calls are for business and government listings, and these listings are much smaller in number, they have started speech-enabling these listings. Nuance has demonstrated a system that aims to support a nationwide DA database of up to 100 million listings. They say that their system should be able to automate about 75 percent of all directory assistance calls.

Another project in the area of EDA is being run by BellSouth IntelliVentures, which has developed a voice-activated yellow pages application. A user dials the service and says he or she is looking for a Chinese restaurant in Gainesville (Florida). This starts a

query process in which the system constructs a series of questions that helps to extract the exact information. The service would ask:

"What area of town are you interested in?" or

"What price range are you looking for?"

Once the system narrows the query down to a set of choices, it begins to read off the names of the restaurants. At this point the service is not able to provide travel directions. But it does give the user the address and phone number and can then actually place the call.

IntelliVentures produces new multimedia products for BellSouth, including The Real Yellow Pages ONLINE and BellSouth AdReach Service. The service began in the Atlanta area and now is free to callers in 404, 678, 770 area codes. The service is positioned as an additional service option for current advertisers. In addition to the enhanced DA, callers get voice-activated information such as weather, stock quotes, business news, lottery results, traffic reports, soap opera updates, horoscopes, and sports reports. Plus callers gain entry to local restaurant guides from 8,000 restaurants in the Atlanta area.

### **5.1.3 Data Devices**

Certainly we are in the midst of a continued evolution in mobile devices, especially a convergence between wireless voice and data devices. For example, Palm PDAs are coming with wireless cards or with built-in wireless capabilities. Sprint PCS offers a Palm Web Connectivity Kit that includes CDMA adapter and CD-ROM software priced at \$59.99 for consumers and \$29.99 for business users. As 2.5G or 3G networks are implemented the new packet networks will enable mobile phones to have larger, more colorful displays. The enhanced services will be connectionless, so they will be "always on."

The introduction of these new products will bring with them, however, issues of customer acceptance. One barrier is usability – entering data via awkward, tiny keyboards and screens is cumbersome on mobile phones. In addition, current services

rely on data speeds in the US that are still pretty slow. Also, not enough Web sites have modified their information content to make it useful to mobile phone users.

The development of useful content is very important. Surely access to stock quotes and sport scores will not sell the wireless Web. But some specific, well thought out applications may. For example, SkillBay.com lets companies use WAP-enabled devices to help them staff and manage IT positions. Registered candidates and vendors can check bid status, schedule interviews and even execute transactions through their mobile devices. Netbank.com enables wireless banking services. Users can check balances and obtain transaction summaries on various accounts.

To make the devices easier to use, companies are adding voice activation. IBM and Nokia, for example, are working to enable speech recognition in wireless phones. As part of the partnership, Nokia licensed IBM's ViaVoice Directory Dialer, which enables automated directory assistance and call routing. Using ViaVoice, the user dials one phone number and, after a prompt, states the name of the person they want to call. The caller is then connected to that person's extension or is provided the requested information such as email, pager or fax number. ViaVoice Directory Dialer is based on IBM patented technologies and is capable of handling in excess of 250,000 names in the directory.

Unified messaging vendors are also optimistic about the impact of new wireless data devices on their market. The addition of WAP on wireless phones is perceived as an important driver for wider residential and small business adoption of unified communications. Many new Web-based services for UM are being launched.

Although market success is not ensured, there are indicators that wireless data use has greater potential. NTT DoCoMo's success in Japan has been tremendous, reaching 15 million subscribers of its wireless Web service in less than 2 years. This is more than 40 percent of its subscriber base. DoCoMo's mobile Internet service is called i-mode. So far what people are doing is downloading news, cartoons, checking bank balances, and sending messages. NTT has invested \$9.8 billion for a 16 percent stake in AT&T Wireless, which intends on offering its version of i-mode, called m-mode in the Spring of

2002. NTT hopes to lead the forefront of adoption of the wireless Web here in the United States.

#### **5.1.4 Voice Commerce**

The next stage in voice portals will be voice commerce -- buying by voice command. Voice portals can bring the same ease of use to e-commerce as they do for information services. One of the key elements to voice commerce success will be security. Fortunately, speaker verification technology is available for secure transactions. Speaker verification can imprint someone's voice the way a fingerprint can identify someone. Each person's voice is as unique as a fingerprint. Voice commerce will make online transactions safer and simpler. Voice surfers will be able to order books from Amazon, groceries from Webvan and office supplies from OfficeMax all without having to talk with anyone. When they first set up their account, users will train the system to recognize their voice and then provide their personal information and credit cards. The Home Shopping Network already has 700,000 customers who have registered to buy products using their voice authentication system. Nuance Communications supplied the software to drive HSN's application.

Another driver in this area will be combining e-commerce with tailored versions of the wireless Web. As this becomes successful, the voice-enabled side will follow suit. One of the major retail groups pushing wireless e-commerce is the hotel industry. Already, hotel reservations can be made through various handheld wireless devices. Bass Hotels & Resorts (e.g. Holiday Inn, Crowne Plaza) offers reservations via Palm VII or Web-enabled cell phones. Choice Hotels (e.g. Qualify Inn, Comfort Inn) makes its reservation system available to Palm users. Hilton, Marriott, and Starwood Hotels are said to have plans to do the same.

Catalog shopping is another potential area. But ordering applications have to be done right, because voice conveys much less information than graphics. For one thing, information must be conveyed one item at a time into a single audio stream. This will

work if the customer is working from a printed catalog (e.g. mail order clothing), or knows the exact product already (e.g. a specific book, CD, or video), or if the products can be described in a sentence or two and there are not many of them (e.g. sports team products). Customer service applications (package tracking, account status, and call centers) are well suited to voice commerce as well, along with financial applications -- banking, stock quotes and trading.

### **5.1.5 Telematics**

Although in its early stages, telematics is expected to expand to a wide range of automobile models (see The PELORUS Group's Wireless Location-Based Markets report). This will be driven by automakers adding telematics units into their new models and aftermarket device makers driving adoption into existing cars. More features and capabilities are being added to the new telematics designs, including speech recognition. Motorola announced an Internet radio prototype called iRadio, which will be sold through automakers. It will let drivers capture music on demand, listen to real time traffic reports, get stock portfolio information, and access email. The product uses a speech recognition interface. Motorola has teamed up with Mercedes-Benz to offer an integrated StarTAC digital phone with voice dialing, using microphones and speakers built into the vehicle. It is already available on some 2000 model Mercedes-Benz cars.

Integrated telematics units can be priced anywhere from \$1,500 to \$2,000 and higher. Units offer a combination of today's voice and data services along with location-based services. They offer navigational systems that employ global positioning systems. Entertainment is also a leading component of the future systems. Passengers will be able to play video games or watch DVD-based movies. Communications, of course is the basic essential. A driver can make hands-free calls by placing a regular cell phone in the cradle of the telematics unit. Voice controls can then be used to initiate and control the call. Internet access can provide email, plus information services such as stock quotes and traffic reports. Safety features are important too. Using voice commands, a 911

operator can be automatically called when an air bag triggers. Vehicles can also monitor and diagnose operational problems automatically. They can present warnings and even suggest appointment times for repairs to take place at a local service center.

Other new applications that will be added to telematics services include smart cards for electronic toll collection. There will also be automatic vehicle identification, weigh-in-motion and automatic border crossing features for commercial vehicles. Plus industry experts see:

- In-vehicle navigation systems;
- In-vehicle information systems;
- Collision avoidance systems;
- Obstacle detection devices;
- Adaptive cruise control; and
- Fleet monitoring systems.

### **5.1.6 Voice over IP (VoIP) and SIP**

IP telephony is the delivery of voice telephone services across the same wires that carry data traffic using Internet networks. IP (Internet Protocol) telephony makes it possible to:

- Retrieve and respond to e-mail via telephone;
- Automatically route calls to pre-set locations without the cost of call forwarding;
- Integrate e-mail contact lists with the telephone;
- Combine a Web site with voice communications in a click-to-- talk call-center application.

Adding to the benefits of IP telephony are substantial cost savings on long distance calls, especially international ones, and the efficiencies that come with an integrated voice/data network. With IP telephony, data or voice transmissions are converted into packets of information that travel across the network - whether it is the public Internet or



intranets or a combination of both. Unlike traditional phone service, those packets do not require a dedicated, open circuit end-to-end, thus enabling more efficient use of bandwidth and network capacity.

Enterprises and service providers are increasingly looking to Internet protocol as a voice transport vehicle. When routing calls over an IP network, analog voice is digitized, compressed and packetized into IP packets. It is then transmitted via a public or private data backbone to its destination. VoIP can be used as an alternative to the public switched telephone network. Some providers offering VoIP services are:

- AT&T Broadband;
- iBasis Inc.;
- ITXC Corp.;
- Level 3 Communications Inc.;
- Net2Phone;
- Tellme Networks Inc.;
- Vonage Holdings Corp.

Voice over IP technology will play a role in speech technology, especially in the voice portal area. Voice portal players have already started looking to offer a VoIP deployment option. This option will reduce operating costs for voice portals. The VoIP connection reduces the cost of the call because it reduces the transfer costs between the carrier and the voice portal company. In addition, next-generation VoIP networks will allow voice to be deployed globally at a very attractive price.

Voice portal firms are making partnerships with VoIP providers. HeyAnita, based in Los Angeles, and Net2Phone announced partnership to provide a voice dialing service. iBasis, a VoIP service firm, has purchased the company, Price Interactive, which is a voice Web application service provider. iBasis plans on offering voice dialing services to its customers using Price Interactive technology. BeVocal will use the iBasis Network to provide European speech application developers with local access to BeVocal speech application development environment using VoIP. Audiopoint will provide ZeroPlus with

50 voice-prompted, commercial-free services where all voice portal and long-distance traffic will be carried over the ZeroPlus United States VoIP network.

Microsoft is emphasizing voice over IP in its strategy. Windows XP operating system supports VoIP's session initiation protocol (SIP). This means conducting phone calls over a Windows XP PC. Microsoft sees a connection between instant messaging and the ability to translate text to speech. TTS IM sessions will be more useful to the user. When on an IM session, one could chat with friends verbally over the Internet rather than type to them. To extend this further, the PC could be used as the platform to make other "calls" reaching people over Internet through VoIP gateways to make less expensive telephone calls.

The major driver of adoption for speech application firms is the ability to reduce or eliminate telephone connection charges to VoiceXML interpreters. Calls made from a PC have no access charges, but VoIP calls from a PSTN phone can only be reduced if VoIP gateways exist in local areas. Thus the key to making VoIP calling cost effective is the upgrading of IP networks for connection to the PSTN networks. Connectivity to PSTN would be accomplished through gateways. VoIP gateways would spread across the globe in local dialing areas would make VoIP calling possible and cheap.

This is happening. Firms such as Delta3, Dialpad, iBasis, ITXC, MediaRing and Net2Phone are involved in gateway distribution. As these companies gain access to VoIP telephony, the phone companies themselves have become interested in providing this alternative. Because they already have local points of presence (POPs), they are better positioned to provide local call access to VoiceXML interpreters.

## **5.2 Next Generation Speech Technologies**

New improvements to speech technology are being introduced each year. Already speech recognition has improved to such a degree that it can recognize most voices and speech patterns. Some of the telephony service providers have a wait and see attitude,

but as speech software improves even the most cautious are beginning to decide for implementation.

Some of the key factors in next generation speech technologies will be:

- Distributed Speech and Multimodal Developments;
- Next Gen Call Centers;
- Natural Language Understanding; and
- Visemes;

### **5.2.1 Distributed Speech and Multimodal Developments**

Two new deployment strategies for speech technology are being widely discussed. One is distributed speech recognition or DSR, which means that speech-enabling functions are distributed between the mobile device itself and the network speech server. With DSR, the embedded device is used in conjunction with server software. Some of the initial audio processing of the voice inputs is done on the device first and then sent along to the network. Part of the drive toward multimodal is SALT. SpeechWorks, along with Cisco, Comverse, Intel, Philips and Microsoft, have formed the SALT (Speech Application Language Tags) Forum, which was formed to accelerate standards for multimodal applications.

The key benefits of DSR are improved accuracy and cost savings. Because some of processing is done on the device first, less must be done in the network. Since some of the first steps are accomplished on the device, less data needs be sent over the communications lines. Another benefit to DSR is personalization, where the system can recognize and take advantage of personal profiles. These personal profiles can be stored and updated by the user on their device and then used by the network to tailor services. This is especially useful with applications related to large distributed sales forces, such as an insurance company that has a large number of reps using smart phones or PDAs. DSR is also especially suited to telematics. One of the key premises of SpeechWorks'

Automotive Devices Initiative is that voice commands are processed on the telematics console first and then passed to the telematics service center.

The other new deployment strategy is called multimodal, which provides the user with dual interface modalities – both the visual and auditory. This can be considered the next stage of speech technology development, because it uses the same DSR structure but adds the visual element. Multimodal provides carriers with the ability to deploy "multimodal" applications, or applications that use both speech recognition and a visual display on a device to create the ideal user interface. This means that users can look at a screen, use the keyboard, and speak commands at the appropriate times. The speech vendors have developed suites of products that enable DSR and multimodal applications for robust modes of interaction with the call center, telephone service or telematics service.

For example, SpeechWorks has laid down a foundation for multimodal and DSR applications with their suite of products. The embedded ASR product, Speech2Go, is complemented by SpeechWorks' ETI-Eloquence software, a small-footprint TTS engine which is available in 13 languages. These two products provide a powerful user experience for a wide variety of applications and can be deployed in combination with SpeechWorks' network-based products to create a seamless user experience through a DSR architecture. The DSR architecture allows speech recognition to happen entirely on the device for local applications (e.g., "tune radio to 105 FM") and splits the processing between the device and servers in the network for network-based applications (e.g., "find Indian restaurants in New York City"). By performing the signal processing phases of speech recognition on the device, DSR systems are able to achieve much higher accuracies than purely network-based systems - especially in noisy mobile environments. SpeechWorks has DSR trials in process with a number of device manufacturers and carriers. To facilitate the adoption of DSR, SpeechWorks chairs the Application Requirements Committee for the ETSI (European Telephony Standards Institute) Aurora committee, which is developing an industry standard for distributed speech recognition.

### 5.2.2 Next Gen Call Centers

Call centers have been renamed contact centers to reflect the many ways that customers, suppliers or employees can interact with an enterprise. Customers are still the most important users of contact centers, but suppliers and employees are becoming a larger proportion of users. Since many channels of communications are available, contact centers really are not physical centers anymore. They have become virtual centers where software now consolidates or blends all contact information coming from various channels.

Some argue whether computer/telephony integration (CTI) is dead. The PELORUS Group thinks that traditional CTI's days are numbered (see The PELORUS Group's report, CTI to CRM). Some people have coined a new term for advanced call center technology - electronic Contact Technology Integration (eCTI). This new term applies to the integration of email, Web/Internet, fax and other media (including speech technology) into the call center.

It is important to emphasize that integration of information with communications networks and customer data is not dead. In fact, many of the same CTI techniques used with voice communications must now be used with all forms of communication. A second new term has been coined, blended media centers, to describe a contact center that manages all the information that relates to all these forms of communications. In other words it could be said that eCTI is not replacing CTI but expanding its application in existing call centers and branching out into new markets where call centers were not originally in place.

Almost everything that can be said about this marketplace today relates to two changes that commenced approximately two years ago and will take probably another two years before most commercial enterprises are fully impacted. First, people are now using multiple communications channels to interact with businesses. Included in these channels are voice, email, fax, Web interaction, and Voice over IP. Voice processing

technology such as interactive voice response (IVR) and speech enabling methods provide easier to use interfaces.

Second, there has been a shift in the way companies view their interaction with their customers and suppliers. Managing this interaction is now called customer relationship management (CRM). The main principle of CRM is that a company must manage the entire relationship that they have with a customer rather than each individual interaction as they occur. Using CTI to provide a screen pop to a call center agent was and is a basic form of CRM. The difference today is that there is a great deal more information to manage, but companies also now have the technology to manage that information. If companies manage that information properly, they will have a more satisfied customer base, which in turn should increase sales to these customers, while saving money on the process of interacting with customers and suppliers.

### **5.2.3 Natural Language Understanding**

Natural Language Understanding is the technique that enables the deployment of systems that use speech in the most natural way. Allowing users to speak in full sentences is the ultimate goal of all speech technology. This goal is still in the future. For now, however, much of NLU is still in the research lab.

The MIT Laboratory for Computer Science, for example, runs its Spoken Language Systems Group (SLS) that conducts research aimed at providing universal access to information via a conversational interface. These interfaces allow users to converse with machines (in much the same way we communicate with one another) in order to create, access, and manage information and to solve problems. Some of the projects that SLS is currently working on include:

JUPITER -- A conversational system that provides up-to-date weather information over the phone. Jupiter knows about 500+ cities worldwide and gets its data from four different Web-based sources.

MERCURY -- A conversational interface that enables browsing of flight schedule and pricing information about via the telephone.

PEGASUS -- A conversational interface that provides flight status (i.e., departure and arrival times) and gate information about currently active flights via the telephone.

VOYAGER -- A conversational system that provides traffic conditions and navigation assistance for the Boston area via the telephone.

ORION -- A conversational agent that performs off-line tasks and contacts the user, at a later pre-negotiated time, to deliver timely information. Some of the applications that SLS has pursued in the past (but is not currently working on) are:

WHEELS -- A conversational system that performs searches through a database of electronic automobile classified ads. Users may conduct searches by telephone or via typed requests, and obtain responses in the form of synthesized speech and displayed tables.

DINEX -- A conversational system that provides information about Boston area restaurants and how to find them. Users can interact with Dinex through speech and by mouse-clicking on displayed lists and maps.

WebGALAXY -- A conversational system operating within a Web browser which combines knowledge sources for air travel, weather information, and local tourist and navigation information for the city of Boston. Users can move easily among domains using verbal or typed input.

As discussed in Section 1.2.2.4,, there are a number of firms in NLU that have developed prototypes, are undergoing trials or have working systems. They include: Ascent Computing, BaBel Technology, Banter, Baobab Natural Dialogs, Brainhat, Cycorp, Delphes Technologies, IBM, Iphrase, Leverance, Nomino, and Simplis.

#### **5.2.4 Visemes**

One of the major new innovations in the speech arena is visemes. The visual analysis, called viseme, is not unlike its sound counterpart called phonemes, which are

the smallest intelligible segment of sound in a word. Viseme is the smallest intelligible segment of a lip gesture, which when put together with other visemes allows the system to recognize significant movements.

Developed by IBM, it is an enhancement to its voice products that includes identification of visual clues. Movements of the lips and mouth are used to help in speech interpretation. Developed by the Human Language Technologies Department at IBM's Thomas J. Watson Research Center, visemes are in beta test with a number of enterprises and will be available in two years. In the beta version, the system uses a microphone, a camera to monitor lip and mouth movement and a set of business rules built into the recognition system. These rules center on social clues, such that if the person is looking away from the camera, the system understands that the person is not talking to it or that if the lips are not moving, the system assumes that any sounds it picks up is extraneous noise. These systems would find application in noisy environments, such as a moving car or on the trading floor of the stock market.

### **5.3 Conclusion**

The most far-reaching visions suggest that speech recognition in the network will cause a revolution. Mike McCue, president of the voice portal Tellme Networks, says that their long-term goal is to create "dial tone 2.0." The replacement for traditional dial tone with universal speech dialing and commands is not that far off. As the phone becomes the access point to the Web, unified messaging and in-car services, voice-activated commands will make the phone a more fluent tool. The telephone already is used to access voice mail, email, buy things online, and search the Web for information. In fact some think that Internet usage will be greater by phone than at the desktop PC in the next decade. Certainly surfing by voice will be a part of this.

AT&T is a believer in voice portals, since it has invested \$60 million in Tellme Networks. They think that over time the new services offered through speech technologies will change the phone business. Instead of picking up a phone and hearing



dialtone, why not hear your personal assistant say, "how can I help you, Dave?" Then the user would respond simply by speaking commands to make calls and find the help and information they need.

Sprint CEO, Bill Esrey, is quoted as saying that voice will become "the Internet appliance of the future." He predicts that voice access to email and voice portals will find great acceptance among all the non-PC users in the world. Voice will be liberated by the new speech technologies to "allow it to penetrate new realms."

The PELORUS Group agrees with the potential that is expressed in these statements, but the path to realizing this is full of twists and turns. This report has summarized the key activities in this market and where the market is headed. Certainly any combination of success among the set of speech technologies: ASR, TTS, SV, or NLU will stimulate adoption of telephony services and provide new revenues. Already speech has created new service revenues and enhanced services such as voice dialing, voice portals, unified messaging and telematics. These will help stimulate wireless usage and support flagging carrier ARPU.

True success will be achieved only if two key factors become true. One factor is technical – that accuracy rates and reliability remain high and even increase. Voice activation must work well in the noisy environments of mobile users and must accommodate to the huge variation of speech accents and languages that one encounters in most countries. Speech technology vendors must continue to dispel the notion that speech technology has yet to be fully proven in the marketplace. It has, but the message has not yet been heard throughout the marketplace.

The second factor that impacts success is the ability to apply speech technology to the range of telephony applications involving the carriers, enterprises and mobile device manufacturers. This means partnerships and adoption by the important application developers, telephony platform vendors and ASRs. Ultimately it seems logical that a combined service that allows the user the greatest flexibility will be the most successful. This means that distributed speech recognition and multimodal deployments will be the

norm. Before this happens, important alliances must take place between the device manufacturers, service providers and speech technology firms.

As the lines of international communication are opened via broadband communications and the Internet, the outlook of many organizations is becoming more global in scope. Barriers such as distance and geography are being eliminated, thus giving companies new business opportunities. Speech recognition companies comprehend immediately the need to provide voice applications throughout the world. Since voice is the universal interface by which man communicates, it only seems logical to build it into the telephony systems.