DMG CONSULTING LLC

The Practical Guide to Speech Recognition

Using Speech Recognition to Decrease Cost and Increase Revenue

By Donna M. Fluss, Principal

DMG Consulting LLC
6 Crestwood Drive
West Orange, NJ 07052
T 973.325.2954
www.dmgconsult.com

Contents

Executive Summary

Speech Recognition is a technology that, when used properly, empowers enterprises to decrease costs and increase revenue. While the technology is ready for prime time, widespread adoption continues to be delayed by limited availability of user interface and scripting expertise. The ROI from Speech Recognition is very compelling for customer-facing, business-to-business and business-to-employee environments. Self-service is increasingly in demand and Speech Recognition provides a proven solution for most business needs.

This paper is a practical guide for business managers considering the benefits, challenges and opportunities for speech recognition. It explores the many applications and uses of speech recognition for enterprises, both in and out of contact centers. The paper explains the technology, presents its value proposition and ROI with case studies, assesses the risks and warns of the pitfalls to avoid. The potential benefits of the current generation of speech recognition applications far outweigh the risks, as long as the project is properly implemented and well managed.

Introduction

Speech recognition technology has a problem - it gets no respect. Speech recognition is compelling and functionally rich technology with many applications in and out of contact centers and delivers a quick 6 to 12 month ROI, yet it's hard to sell. Adding to the paradox, it's one of the few, if not the only self-service technology that should be intuitive for customers, even without training, as it uses the most ubiquitous form of communication - speech. But companies remain reluctant to make this investment.

Customer service is going from bad to worse as the weak economy pressures enterprises to reduce costs. Conflicting with, or maybe in response to this reality are recent studies showing that customer service is the most important initiative for more than 65% of enterprises in 2002. But investments are not keeping up with stated priorities; enterprises are squeezing customer service budgets and the quality of service is declining. The service market is demanding technology that simultaneously reduces cost and improves quality and speech recognition has what it takes. Speech recognition technology is very compelling, particularly in a challenged economy that purports to emphasize service quality, so it's counterintuitive that vendors in the speech market are continuing to struggle and most have not yet achieved profitability.

Worldwide investments in speech recognition technology are $650 million today and are expected to grow to $5.6 billion in 2006, according to Datamonitor. The market is ripe for speech recognition and the value proposition and economic benefits are clear, yet these products are not realizing their sales potential. Speech recognition should be as common as Interactive Voice Response (IVR) systems and in some companies should actually replace IVR applications, but it's just not happening on a large scale.

The adoption rate of speech recognition in Customer Relationship Management (CRM) and call center environments, its primary market for the last 20-plus years, has been painfully slow. The acceptance of speech and voice technologies in non-CRM areas during the last two years, however, is exceeding its rate of implementation in call centers and will drive its success in call centers and CRM during the next five years.

The Benefits

The benefits of a properly implemented speech recognition application are huge. An effective speech recognition project will realize a payback in its first year. The catch (and there always is one) is that speech applications need to be designed from the user's perspective and the user interface (UI) has to be customized for its audience. Bluntly put, an application designed for New Yorkers is not going to work well in Texas.

Speech recognition is generally viewed as a contact center productivity tool because that's where the financial payback has been the greatest, but it has already proven effective at generating revenue for companies in and out of contact centers. On the productivity side, speech recognition reduces calls to agents, shortens call lengths, reduces call hold times and decreases call abandonment rates. Speech recognition technology allows enterprises to generate revenue both by automating activities that

previously required agent assistance and by providing new services, including voice portals, voice activated dialing and e-mail reading.

During the recession, many call centers have downsized, operating with minimal staff, which has resulted in reduced service levels. When the economy recovers and volumes increase in call centers either new staff will have to be added quickly or there will have to be investment in new self-service technologies that allow customers to help themselves.

Speech recognition can benefit enterprises whether or not they have an existing touch-tone IVR system. If the call volume is high enough, even a company that already realizes an 80% touch-tone IVR utilization rate may get a great payback from speech recognition. According to an April 2001 report from Giga, adding speech recognition capabilities to an existing touch-tone IVR platform "increase usage by 20% to 60%." See Table One below.

Table One: Increasing Self-Help with Speech Recognition

Company	Amtrak Train Status (2)	Dreyfus (1)	Merrill Lynch (1)	Michigan Department of Treasury (2)
Touch-tone IVR usage rate	42%	45%	82%	10%
Usage rate after speech recognition implementation	70%	63%	90%	98% (tax refund status calls)
Call volume	2.8 million/year	12,000/day	50 million/year	2 million/year
Annual Savings	~$1.2 million	~$1 million	$6.3 million	$ 2 million

Sources: (1): Nuance, (2): SpeechWorks

Speech Recognition History

Speech recognition technology has a great history. In 1997, after approximately 20 years of development, the accuracy level of speaker independent directed speech recognition was greater than 95%, giving enterprises the confidence to use it in call centers. (At a 95% accuracy level, speech recognition is as accurate as many call center representatives and often more polite!) Unfortunately for speech recognition vendors, this was also the time the Internet took off and too many enterprises decided to invest in web initiatives instead of speech recognition, even though the potential return from speech recognition was higher. We all know what happened with the web investments once reality hit, but the economic slowdown followed by the recession

continue to put great pressure on enterprises to minimize new technology investments. Speech recognition is again a victim when it should be a must-have.

Speech Recognition technology is ready for prime time. It has confronted and addressed many technology challenges in the last five years. System enhancements include:

1. Achieving an accuracy threshold of 96% to 98% on speaker independent directed speech recognition,
2. Support of vocabularies in excess of 1 million words,
3. Adding "barge in" capability to improve system usability,
4. Reduction in computing needs for speech applications,
5. 99% recognition accuracy using voice verification in conjunction with speech recognition,
6. Development of natural language-like interfaces that improve ease of use, and
7. Introduction of vendor independent VoiceXML development standards.

Although speech recognition technology is viable, there are few primary providers due to the high start-up investments. The leading vendors in speech and voice technologies in the United States are Nuance, Speechworks, and IBM. Outside the US, particularly in Europe, Philips is a contender but they have not been successful in the US.

Market Penetration Obstacles

There are three major obstacles slowing the penetration of speech recognition systems in service environments: proprietary coding languages, the difficulty in developing effective customer interfaces and the cost and time required to build speech applications.

The first challenge, proprietary coding languages, is being addressed with VoiceXML. VoiceXML, introduced in late 1998, is a non-propriety, open development standard for voice applications, and is changing the perception of how difficult it is to code voice applications. All four leading speech vendors, Nuance, SpeechWorks, IBM and Philips support VoiceXML standards and participate in industry committees to enhance this language. VoiceXML is still maturing, so it's three to five years away from being flexible and complete enough to address the more sophisticated speech applications. Developers still need to use proprietary development code for advanced speech applications, but VoiceXML is fine for basic functionality.

The second major obstacle is the difficulty in designing effective scripts and user interfaces -how the speech recognition system communicates and interacts with system users (internal or external customers) and makes the difference between customer acceptance and rejection of a speech application. The lack of true speech recognition scripting and customer interface domain experts (speech scientists) is exacerbating this problem, as are claims of expertise by companies that clearly lack it. Many speech recognition resellers and consulting firms purport to have the resources to design user-friendly speech recognition applications but experience keeps pushing

prospects and users back to the primary vendors whose experienced speech scientists have designed more than five speech systems. When hiring external scripting and customer interface experts, avoid a false start by verifying that the company you are working with has experienced application coders, scriptwriters and speech scientists. Look for people who have built five or more applications and are familiar with your industry.

The third impediment to investments in speech recognition technologies is the perception that it's very expensive to build speech applications (see note 1) and that investments in other technologies and applications have higher and quicker returns for companies. Managers are having a difficult time justifying investments in speech applications, even though the payback from speech applications in a service environment can be as short as six to nine months, even for a complicated application. It's possible that e-mail response management applications that are implemented with the proper procedures and processes could yield a higher ROI and it's likely that once CRM analytics software delivers on its promise to increase revenue it will have a higher ROI. But this isn't likely to happen anytime soon, so for the foreseeable future, speech recognition appears to be the clear winner.

If speech recognition tracks in a similar pattern to the IVR market, common acceptance in corporations won't happen until enterprises are able to own their development and interface resources. (As these applications change frequently and sometimes even daily, it's important to be close to the resources). VoiceXML addresses the coding issue but not the script and interface challenge. True speech recognition scripting and customer interface experts are rare, many have PhDs in speech and related areas and they are very expensive. At the current pace it will be three to five years before there is a market of qualified speech scientists.

Factors Driving Acceptance and Adoption

When the economy begins to recover, speech recognition will be in great demand in contact centers to automate the handling of calls and empower customers to do more for themselves. Service environments will be able to "do more with less," employing fewer new agents to handle increasing call volumes. But success with speech recognition is only going to happen if implementations are made easier with simpler and more standardized development tools (voice XML) and better scripting and UI resources.

Two trends may speed up the acceptance of speech applications in service environments, the increasing use of voice and speech technologies for many functions, from hands free cellular phones to talking cars, and voice Application Service Providers (ASPs) and packaged application providers.

Voice ASPs and packaged application providers (see note 2) are trying to address speech recognition's current limitations with out-of-box, verticalized systems that do not require a great deal of customization. Unfortunately, many of the enterprises looking for short cuts to entry in the speech market are not willing to compromise

their service differentiators; they don't want to use the same application as their competitors.

The Voice ASPs and packaged solution providers are on the right track by delivering verticalized applications but need to modify their business models. Out-of-box applications are a good starting point but Voice ASPs need to provide application development and customer interface expertise at reasonable prices.

Growing Demand for Speech Applications Outside Call Centers

Until 2000, the predominant use of speech recognition was in call centers in conjunction with IVR applications. As speech remains the most ubiquitous form of communication and is unlikely to be replaced by the web in the next five years, there's been a growing demand for speech-empowered applications from many industries hoping to reduce their costs while improving the quality of service. Uses include:

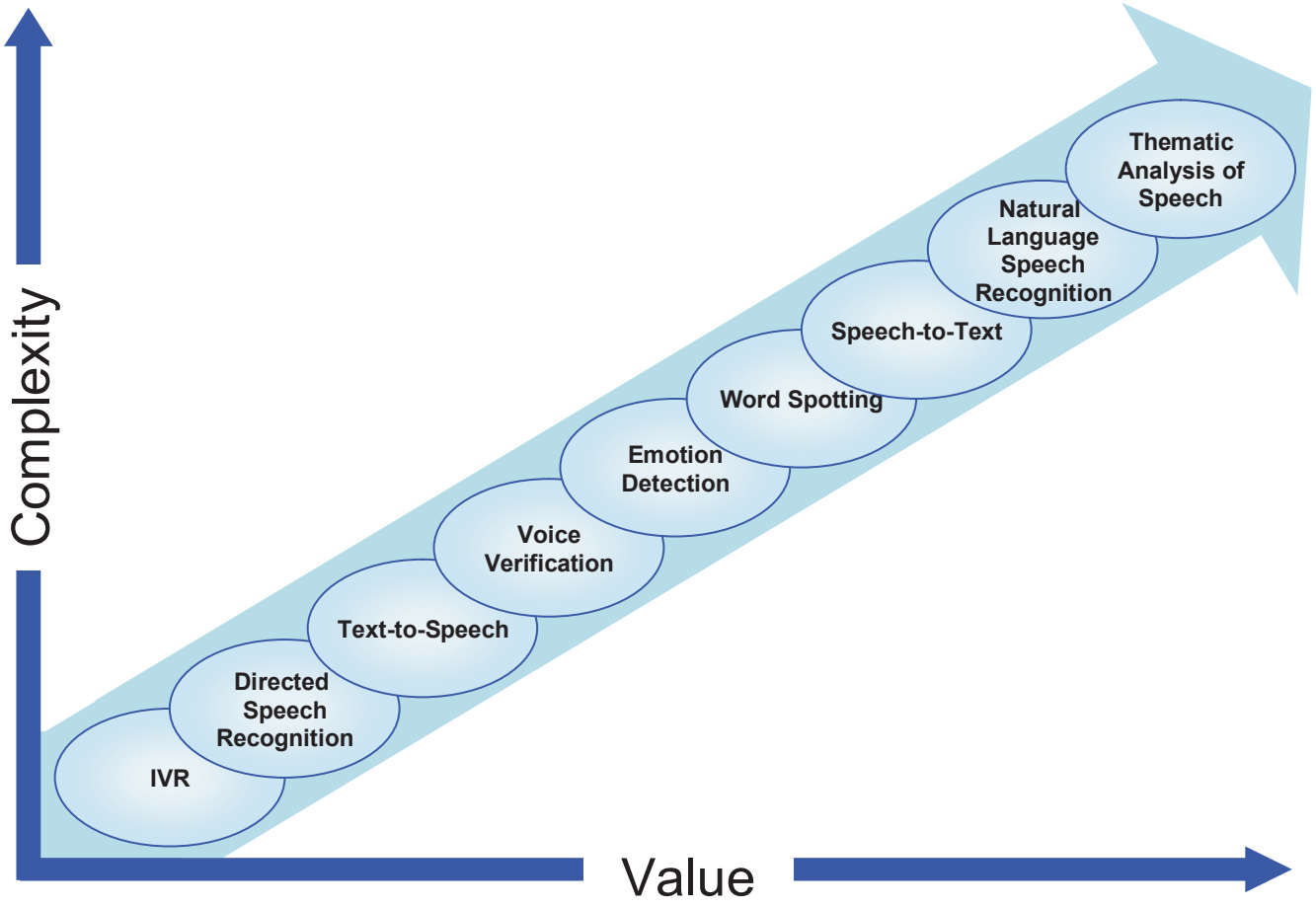
- **Content**
  stock quotes, sports, news, weather, and horoscopes;
- **Retail**
  placing orders, checking prices and availability of stock items and locating stores and addresses;
- **Telecos**
  voice activated dialing, information portals, phone-based e-mail readers; Directory assistance - speech-enabled directory assistance to replace overworked and often unpleasant phone operators who shouldn't be serving the public;
- **Advertising**
  permission-based advertisements to fill time spent waiting for service agents;
- **Government**
  applying for loans, checking the status of filed documents, locating the closest post office, directions;
- **Transportation**
  providing train and airplane schedules, booking travel reservations, changing or canceling reservations, checking hotel rates and availability, checking status of loyalty programs;
- **Field Service**
  checking the status of parts, scheduling service visits, placing orders for repairs and parts;
- **Entertainment**
  identifying location of movies or shows, dinner reservations, ticket purchases, directions;
- **Credit Card**
  authenticating the caller, obtaining account balances and available credit, making payments, transferring funds, requesting copy of statement, activating account, reporting lost or stolen cards, new marketing promotions;
- **Call Center**

product and service information, making payments, placing orders, reviewing order status, updating personal information, applying for credit, applying for jobs, locating stores, paying bills, pricing information, marketing promotions;

- **Parcel Services**
  requesting a pickup, package status, office locations and directions, rate calculations;
- **Automotive**
  basic car instructions;
- **Energy**
  meter tracking, reporting gas leaks, making payments, payment due dates.

The underlying technology enabling speech recognition is maturing, opening up new opportunities and expanding its potential uses throughout enterprises. For example, speaker independent Speech-to-Text (STT) technologies are developing and are expected to be viable within the next three to five years. Speaker independent STT applications have huge potential within the CRM and call center markets as they open customer conversations to thematic analysis and analytics, which will allow enterprises to mine these conversations for new revenue opportunities (privacy issues aside). Marketing organizations are anxious to capture and leverage the information freely shared by customers in call centers but do not yet have automated tools to fully understand customers' intents. Speaker independent STT gives them these tools.

Figure One: Speech Technologies



Future Applications

There are many non-call center uses for speech recognition in an increasing number of industries, including:

- **Library**
  placing books on hold, ordering books, checking book status, paying late fees;
- **Education**
  requesting college applications, registering for college courses, setting up appointments for teacher meetings, ordering class materials;
- **Medical**
  emergency doctor calls, ordering medicines, scheduling hospital appointments, scheduling the operating room, referrals;
- **IRS**
  ordering tax forms, tax questions, checking status of tax refunds;
- **Insurance**
  checking for coverage for medical procedures, checking status of insurance payments, determining if a doctor is on a plan;
- **Airlines**
  ordering special meals, seat selection;
- **Auction Houses**
  answering questions about sales, helping prospect locate specific items for sales, accepting "phone bids," payments;
- **Motor Vehicles**
  scheduling driver tests, license replacements and renewals, registering cars, license plate orders and renewals;
- **Sports Teams**
  purchasing tickets, selecting seats, reselling tickets, buying paraphernalia.

These are just a few potential future applications of speech recognition, but the list displays the flexibility of this technology. Directed speech recognition is very flexible and allows enterprises to automate many activities currently being handled by expensive service agents. Automating basic functions will reduce operating expenses and improve quality as studies show that service representatives often get bored handling mundane activities and do a better job dealing with more complex issues.

Return on Investment

Before moving ahead with an investment, enterprises must analyze their current operating environment and conduct a Return on Investment (ROI) analysis to determine if a speech system will enhance customer self-help functionality and off load more calls from agents. Below is an ROI analysis for an enterprise that receives one million calls per month and had a 62% touch-tone IVR acceptance rate prior to its speech implementation. This company carefully studied how its customer base used its touch-tone based IVR and identified new self-service options and service enhancements for its existing touch-tone IVR platform that could be made possible only with the addition of speech recognition. Adding speech recognition capabilities increased the self-service rate to 75% and had a pay back of less than one year. The net present value (NPV) and internal rate of return (IRR) for this project, over a three-year period, were $ 18,792,795 and 640% respectively.

Table Two: Speech Recognition ROI Calculator Model

Monthly Volume	
Total incoming calls	1,000,000
Fully Automated Sessions	
Touch-tone IVR only	62%
Touch-tone / Speech System	75%

Expenses [1]	
One Time Investments	
IVR Ports (hardware/software) [2]	$ 480,000
Speech Ports (software) [3]	$ 424,000
Development, Implementation and Integration	$ 400,000
Annual Costs	
Hardware Maintenance (18%)	$ 86,400
Software Maintenance (18%)	$ 76,320
In-house Support (1/2 FTE) [4]	$ 50,000

Benefits	
Projected Monthly Savings	$ 715,000
Projected Annual Savings	$ 8,580,000
Projected Annual Costs	$ 212,720
Net Annualized Savings	$ 8,367,280

Return on Investment	
Pay-back Period (months)	1.8
Internal Rate of Return	640%
Net Present Value (3 years @12%)	$ 18,792,795

Assumption	
Cost Per Agent-handled Phone Call [5]	$5.50

(1) The model assumes that back-end systems and processing costs are the same regardless of how calls are handled.
(2) The model assumes that that there will be an equal number of IVR touch-tone and speech ports. 120 incremental IVR touch-tone ports, at a cost of $4,000 per port, were added to accommodate the increased use of the system.
(3) 265 speech ports are required to accommodate the busy hour of 1 million calls/month. The cost per speech port is $1600.
(4) The model assumes that adding speech recognition capability to an existing touch tone system will require incremental in-house technical support. The cost of each support person is assumed to $100,000 per year.
(5) Gartner Measurement estimates that the cost per agent-handled phone inquiry is $5.50.

Notes:
1. Call categories that can be automated with speech recognition include: Information requests, transactions, calls that involve multiple levels of verification or long verification numbers, alphanumeric or auto attendant.
2. Examples: balance requests, call routing, on-line ordering of products or services, travel reservations, seat requests and confirmations, stock trades and confirmations, fund transfers, verification and other banking activities, address changes, library book orders, requests for tax or insurance forms, newspaper deliveries, etc…

Investment Guidelines

After years of relatively free spending in the late 1990s, many companies have implemented strict investment guidelines; the office of the CFO is watching investments very carefully. Investments have to have measurable benefits based on hard dollar savings in order to be approved. Cost centers, which include the majority of call centers, can justify investments in productivity improvement, cost reductions and cost avoidance. Profit centers, which include sales and marketing organizations, can justify investments based on productivity improvements and increased revenue. Because it's no longer acceptable to base spending on soft dollars, a cost center cannot justify investments based on increased sales leads or quality improvements (although "soft" benefits often contribute to hard dollar savings). (See Table Three below). As a practical matter, this means that speech recognition projects that result in call reductions and a 6 to 12 month ROI will be approved.

Enterprises and vendors will have one chance to do it right. No excuses will be accepted for project failure. If the project fails, project sponsors will be at risk of losing their jobs.

Table Three: Typical IT Investment Approval Criteria

Organization	What will be approved	Examples
Cost centers	•Productivity improvements •Cost reductions •Cost avoidance	Reduction in: •agents •supervisors •telecom costs •real estate •training expenses •hiring costs •quality assurance costs
Profit centers	•Productivity improvements •Cost reductions •Cost avoidance •Increased revenue	Increase in: •marketing campaign response rates •revenue/sales •new customers •customer retention Reduction in: •acquisition costs •marketing costs •sales costs

The Implementation Matters!

Speech vendors are now pushing enterprises to eliminate their touch-tone IVR platforms. By all accounts and studies, touch-tone IVR is the most hated form of customer service, but if the application satisfies a large percentage of your customer base - meaning, your customers use the touch-tone IVR platform and do not opt out to agents - then replacing it with an unproven speech recognition application is not recommended as it will cause a service disruption. Instead, speech should be added to the existing application. Speech empowers the IVR platform and adds new features not previously possible with touch-tone. Over time, it may be a good idea to replace touch-tone IVR applications with speech, but not before your customers have a chance to get used to the new application. Additionally, while touch-tone IVR annoys customers, it may very well be the least expensive form of customer service. Customers accustomed to obtaining their account balance on a touch-tone IVR platform can do so in seconds.

Customers don't like surprises or changes in service, although they do enjoy optional service enhancements that are not forced upon them. When properly implemented and introduced to customers, speech recognition applications will improve productivity, quality and overall customer satisfaction.

Compelling Cases

The ROI for speech is very compelling, as the following case studies illustrate. United Airlines and United Parcel Service (UPS) started with small speech recognition applications and added new functionality when the initial applications proved overwhelmingly successful. Both companies realized quick payback, reduced operating expenses and improved quality and customer satisfaction.

In 1998, United Airlines, the second largest air carrier in the world, wanted to reduce the cost of processing 1.5 million employee calls to its reservation representatives requesting complimentary travel. Working with SpeechWorks, United Airlines implemented a speech-empowered employee reservation system within four months. United's employees were very pleased with the self-service speech system since it was personalized and confidential. The employee reservation system paid for itself quickly and freed United's reservation specialists to handle revenue-generating customers.

In August 1999, United rolled out the industry's first delayed baggage speech recognition module. Using a speech recognition system to handle a sensitive issue like delayed baggage was challenging, but the automated service has been well received by customers and employees.

Encouraged by the success of the employee reservation system and delayed baggage application, United quickly proceeded to build a flight information speech recognition system for its customers in five months. This module answers questions about United's 1700 daily flights, handling 40,000 to 60,000 calls daily and 200,000 calls at peak. Off loading basic informational inquiries to the speech recognition system frees United

service agents to handle more valuable revenue-generating calls.

United is continuing to bring to market innovative speech recognition applications that benefit its customers and shareholders and is considering new opportunities to use speech recognition within its call centers. United knows that speech recognition isn't for everyone, so it allows its customers to opt in or out of the system. But customers who do use the speech-empowered self-service options find them convenient and satisfying.

United does not publicly share its ROI data but its continuous investments in speech recognition applications reflect their huge success.

United Parcel Service, the largest express carrier and package delivery company, with 2001 annual revenues of $30.6 billion, delivers more than 13 million packages and documents per day for 1.8 million shipping customers and 6.1 million recipients worldwide. In 1997, UPS was faced with the challenge of distributing more than 265 million packages in a single month. Dedicated to providing outstanding customer service with minimal or no hold time, UPS knew it couldn't staff up its already large work force of 6000 operators quickly enough to handle the anticipated call volume without a decline in service quality. Instead, UPS decided to implement a speech recognition application to handle inquiries about package status. Working with Nuance, the system was implemented within four months, in time to handle peak holiday call volume, and paid for itself within three months. The tracking application typically handles 120,000 calls a day but can handle a peak volume of 936 simultaneous callers and 240,000 daily calls.

Since 1997, UPS has continued to introduce innovative and helpful speech recognition applications for its customers. In 1998, UPS deployed a package pickup module that has shaved two minutes off pick-up requests and has also automated the delivery notice follow-up process to enable customers to locate packages that could not otherwise be delivered. In 2001, UPS deployed a cost calculator module that allows customers to determine the price of package deliveries. As valuable as these innovations have been for saving money and speeding up service delivery, UPS's customers are also " comfortable with these friendly speech applications because they are in control, " states Joan Madden, Project Manager. "They know our operators are available to assist them 24/7 but are increasingly using our speech applications because they meet their needs."

Conclusion

Within five years, speech recognition technology will become so pervasive in our daily lives that service environments lacking this technology will be considered inferior. Speech Recognition technology is now offers viable productivity and revenue-generating opportunities for many companies but the remaining challenge is the limited availability of experienced user interface and scripting specialists. It's expected to take another three years before speech recognition scripting and user interface experts are readily available but enterprises cannot afford to wait. It's time to invest in speech

recognition, even if your script isn't perfect at first. Just as the market did with IVR applications in the 1980s, carefully monitor new speech recognition applications and be prepared to make frequent scripting changes. As scripts and interfaces are enhanced, your company will develop the necessary scripting and user interface expertise.

No initiative is without risk but the potential benefits more than justify serious consideration of an investment in a well-designed and implemented speech recognition application.

Notes

Note 1: Speech Recognition Cost Estimates. See Table Four below.

Table Four: Speech Recognition Cost Estimates

Category	Cost	Comments
Cost per speech port	$200, for a very basic system, to $1600, for advanced natural language recognition capabilities	Expect to have to purchase a minimum of 12 ports
Application Development	$35,000 for a basic auto attendant application to more $500,000 for a sophisticated application	Costs can be reduced by using an out-of-box application from a Voice ASP
Cost per IVR port	$800 to $1500, depending on the size of the system	In service environments, the majority of speech recognition applications are implemented in conjunction with an IVR

Note 2: Packaged Application Providers and Application Service Providers (ASPs):

Packaged Application Providers:

BeVocal, Datria Systems, Fonelet, General Magic, Informatica, Informio, JustTalk, Mitel Corporation, Motorola, NetByTel, OnMobile, Openwave Systems, Siebel Systems, Sierra Atlantic, Skyflow, Telera, TellMe Networks, Incell, TuVox, VeCommerce, Webversa

Application Service Providers:

Appriss, Audium, BellSouth, BeVocal, Blue Wireless, Call Interactive, Centerpost Corporation, Covigo, Cusrious Networks, Diagenix Corporation, General Magic, HeyAnita, High Acre, iBasis Speech Solutions, iHello, Informio, Interactive Telesis, IntoVoice, Inzigo, Loquendo, Magic Phone, Mitercom, Motorola, NetByTel, Omron Corporation, SkyFlow, SpeechHost, Tellme, Telera, Time iCR, Vail Systems, ViaFone, Voxeo Corporation, Webley Systems, West Corporation, Wirenix, WorkForce Technologies.

Note: This is a partial list of the voice packaged application providers and ASPs.

About DMG Consulting

About the Author

Donna is the Founder and Principal of DMG Consulting. She is a contact center and analytics expert with over 19 years of strategic and operational experience in Fortune 500 companies and technology enterprises.

Donna is a former Vice President and Research Director in Gartner's CRM Practice and a recognized visionary in all areas of customer relationships, service and support. Prior to joining Gartner, Donna built a reputation as an expert in business transformation and mergers working as a Vice President at Chase/Chemical Bank. She is a well-known industry analyst whose ability to identify trends, directions and needs has made her a leading advisor to companies and vendors throughout the world. Donna is a highly respected writer and speaker. She is a columnist for Call Center Magazine and a frequent contributor to numerous publications, including ICCM Weekly and Customer Interface Magazine.

About DMG Consulting

DMG Consulting is the premier advisor in the areas of customer relationships, service and support with over 3000 consulting engagements to Fortune 500 companies and technology vendors. Our methodology, which includes in-depth analysis of strategy, operations and technology, has a direct impact on revenue and expenses and delivers measurable results.

Contact Information

6 Crestwood Drive
West Orange, NJ 07052

tel 973.325.2954
fax 973.325.2071
e-mail info@dmgconsult.com
url www.dmgconsult.com

Donna Fluss donna.fluss@dmgconsult.com