Voice-Driven Enterprise Applications

September 13, 2000

Proceedings

The Silicon Valley World Internet Center and its programs are supported by:

• Amdocs, Inc. • Deutsche Telekom • Hewlett-Packard Company

• IBM Corporation • Fujitsu • SAP • Sun Microsystems • Metiom

Executive Summary

Speech understanding and speech synthesis have been around as a major research initiative for 30 years. Recently, systems have become more accurate, more pleasant to listen to, and less expensive to implement. Speech understanding has been successfully deployed in certain, contained applications.

Since it is expected that most people accessing the Internet globally will do so through wireless phones and personal digital assistants (PDAs), many believe that voice technologies are inevitable.

Additionally, voice recognition technologies are being quickly aligned for large-scale systems deployment across enterprises.

Definitionally, these applications consist of a presentation layer using speech input to interact with information systems across and within an enterprise. That layer may include applications that provide non-voice driven outputs and are delivered across different devices (phones, computers, PDAs, etc.).

This Think Tank Session focused on how these technologies could be applied and utilized for cross-enterprise applications to enhance effectiveness and productivity of business applications, and where lucrative markets for those creating voice-only solutions may be.

Presenters included Sunil Soares, Program Director of Product Management for IBM's Voice Systems; Bruce Grant, Key Practice Director of Advanced Technologies for Luminant Worldwide Corporation; Bryan Mekechuk, Chief Technology Officer & Co-Founder for Voci Corporation; and Linda Cliff, CAAPS Business Project Manager & Senior Consultant for Kaiser-Permanente's Enterprise Integration Group.

There was discussion on which applications will be most relevant, including speech and language recognition and understanding; text to speech and media playback; and call control and transfer systems. Current uses by businesses range from systems which cut costs within call centers, to enhancing customer satisfaction and company image; and improved productivity of employees.

Purpose and Scope of Think Tank Session

The purpose and mission of the session was to examine voice-driven enterprise applications and their viability in the marketplace in the next 12 to 18 months.

In discussing current markets, participants identified the following issues:

- Where the current market is for enterprise systems
- Which applications are most effective at enabling broad enterprise usage
- The role of Voice XML
- Assessments of key usage areas
- Profitability of the market
- Needs being met in the market

Voice Interaction with Devices

IBM's Sunil Soares, Program Director of Product Management for IBM Voice Systems, spoke at the Think Tank Session. Mr. Soares' talk touched on how IBM has invested in and developed an assortment of voice-related systems, including speech recognition, voice server, unified messaging, and a message center -- a complete solution set.

"The Mobile Internet is driving new applications," said Soares. "As my boss is fond of saying, 'Our fingers aren't getting smaller, but the devices are.' This points us towards voice interaction with devices.

"IBM is about e-business. Voice is the next phase in e-business. Important standards are continuing with HTML in 1994, IVR directed dialog in 1999, Bluetooth in 2000, VoiceXML, wireless application protocol (WAP), VoiceTIMES, natural language and multi-modal browsing.

"Traditionally, applications for businesses supported call-center infrastructure. Now, there's a movement to

voice for e-business, WAP (Wireless Application Protocol), and self service. VoiceXML will help to integrate many of these new opportunities.

"E-business consists of a data network (the Web) and a voice component with self service modes and mail modes. Ninety percent of people don't complete a transaction on the web because they don't have a live agent to get them through the buying process.

"Only IBM provides an end-to-end voice-enabled eRM solution with data, voice and ASP. We use JAVATM (previous is a trademark symbol) beans for WebSPhere Voice Middleware, and enable multi-channel access.

"eRelationship management is important. IBM understands the value of the voice promise. Speech, telephony, and web all must work together.

"We are looking to write the business application once and access it across different environments (voice, visual, etc.), where each presentation layer is different. Back-end logic can be the same in each layer. When a consumer wants to conduct a transaction, the data base, web servers, etc., can be the same, regardless of variety in the modes."

A General Overview of Where the Markets Are Going

Luminant Worldwide Corporation's Bruce Grant, Key Practice Director for Advanced Technologies, provided quick bullet items about his views and experiences on voice applications.

"Luminant is a systems integrator that supports global 1000 clients," said Grant. "We have a pervasive business where we are working on wireless, web, and voice solutions. We sell high in the organization and early, largely at a board or CEO level.

"Voice recognition is coming. In terms of how businesses are investing, the wireless Internet is here, while voice is coming later. Most businesses want to look at it next year. The Gartner Group said that voice would be the way to access the Internet in 2001.

"We think that portals should be led by merchants, and that voice will allow for quick-hit interactions.

"The German government is going to ban phone usage in automobiles except for voice-enabled phones."

So why is voice technology taking off now? According to Grant, we now have faster computers that process more information. Many more people have cell phones than PCs, so the way people access the Internet is changing. VXML will provide a standards-based technology, and there is improved compressor and decompressor software.

Since there are risks in going to market too early or waiting too long, companies should invest in small step applications and hone in on those that are most applicable and effective. It has been stated that 27 percent of the population would have problems with voice because of dialects, inflection, etc.

Luminent suggests companies limit voice input to short lists, high proposition applications. Also look for ways to write once and read everywhere. Base the design of user interfaces on focus groups. Businesses have to work to understand the customer. There can be false starts.

"We think more personalized applications are coming. Voice is a key area for personalization. Carriers will lose out in this. Online wallets will come back (purchase from voice interface). We think portals will flounder in the short term, but will come back. Personalized information will gain rapidly in momentum. PDAs, phones and multimodal will be key. Voice server providers will take off."

Looking at the Standards for the Industry

Bryan Mekechuk of Voci moderated a discussion on which technology and applications will drive growth in the market.

The session looked at the standards that will be important, what is being used in the market today and where we might be headed. Current usage areas are in dictation, web browsing, speech-to-text and command and control.

Michael Cohen, Vice President of Dialog R&D for Nuance Communicationswas asked for his input.

"Nuance builds speech recognition, language understanding and speaker verification technology," said Cohen. "We were started in 1994. Examples of call center applications we have deployed include brokerage services (a customer can get quotes and execute trades), travel systems (schedules, changing reservations, driving directions), and banking. These well-defined domain areas for information access are easily handled with current technology.

"We see that callers who previously transferred to live operators will talk to voice recognition systems, leading to agent replacement and reduce costs."

"Return on investment is key; we have seen a number of clients with ROIs within 2 to 3 months. We see that callers who previously transferred to live operators will talk to voice recognition systems, leading to agent replacement and reduce costs."

The group was asked to define the main functional areas for voice recognition technology. The following were the most popular and agreed-upon responses:

1) Language understanding, speech recognition and verification
2) Text-to-speech, media playback or delivery/audio output
3) Call Transfer (call control and data transfer),

In narrowing down to three areas, speaker verification and identification, command and control, and web access were not addressed.

According to the group's collective knowledge, the vast majority of our systems today are really call center applications, and are not web based. These systems provide an alternative from logging into the web and they make some transactions easily and quickly. At this point one can give directed transactions, but not yet free-form.

Command and control applications are available. It is possible to use a desktop which is controlled by voice. The software's been out for over three years and is extremely accurate.

Constrained dialogue and open dialogue involve the process of assigning meaning to a word stream. After the recognition system knows the words, it has to assign meaning and reply with proper code. Natural language assigns meaning to a word stream.

The application areas listed by the group are shown in Table 1.

Voci's Beta Test with Nova Scotia Power Company

VOCI Corporation's Bryan Mekechuk, spoke to the group about their beta test with Nova Scotia Power Company.

"We ran a pilot program with Nova Scotia Power Company," said Mekechuk. "We were an ASP for voice recognition, and we charged on a per-call basis. This utility experienced infrequent callers and there was little incentive to offer self service.

"A utility is unlike a brokerage firm where there could be repeat callers who are highly motivated to use a self-service system. Those people tend to be better educated, high net worth people, more familiar with self-service.

"Nova Scotia has 130 people in the call center, and the centers are expensive to run. This self service is complementary to the Internet. It helps elderly customers and others use rotary phones. Customers like the voice sys-

Table 1: Technology Issues Applied to Application Area

Standards / Technologies	Speech Recognition	Audio Output	Call Transfer
VoiceXML	• Faster to develop than modularized dialogue components • Limited complixity • Limited natural language understanding • Easier to learn • Lacking re-usable framework	• Lack of re-useable industry framework • Problem with audio output subsystems • Expensive to create commercial-grade solutions	• Lack of re-useable industry framework • Limited support of voice telephony • No support for data transfer • Requires working with CTI APIsto add call control (VXML could do this better with good standards)
CT Access Applications	N/A	N/A	• No standards by telephony vendors • Long development time • Integration problems with partners and VXML
Modularized Dialogue Components	• Re-useable, customizable ability to do intricate code and business logic • More complicated to learn and develop • User interface is available (help messages) and capsulized • Has to interface with proprietary systems • Client versus Server • Components can be used by VXML	• Same as Modululariza tion Dialogue Components / speech recognitions issues, as well as the VoiceSML / audio output issues	

tems because of ease of use, perception of fast response, greater sense of control, and the impression that the company is looking to the future and making it better for the customers.

"We had a 92-percent rate of processed call completion. Eight to 10 percent of calls were completed with self service versus four to six percent for touch tone."

The benefits to Nova Scotia Power were to build the business, develop customer loyalty and develop employee commitment.

"We charged on a per-call processed basis.

"Now we are working for catalogue companies who love the voice application. Our equipment is sitting on their premises. With catalogue companies, you phone their location and requests get routed to them sitting in a data center.

"The hardware is NT, and carrier-grade Dell boxes. We use Natural Microsystems, and partner with Edify."

Potential Challenges or Pitfalls of Voice Systems

In the course of the Think Tank Session, the group brainstormed a number of issues. The following list identifies some of the challenges or potential pitfalls of having voice systems.

What would hold back the advancement of these opportunities?

- *Long distance and multiple phone, cell and Internet charges.* It is not cost effective.
- *Interface is not there.* It still needs to become more effective to use.
- *We read information differently than we hear it*, so writing once and using everywhere might not work.
- *Human factors/useability*. We can skip and scan on a screen instead of listening in a string. What are the types of applications that can work? Hearing versus reading.
- *Attachments in the email*. Voice cannot access the attachment, depending on the application.
- *Free form, natural language recognition.*
- *Synchronization and Contact Manager* – Sync issues with Palm and others (though Speechworks and Palm are going to do synch). Expectation is to be able to use applications from a car or from the office. Different environments need to sync with one another.
- *Simple context for the systems*. How do you express urgency or personal preferences, if you have disabilities? Context in audio may not be as good as a PowerPoint presentation that visualizes a point. Some messages need to be input as a chunk instead of a string of inputs.

The group's thoughts on design and implementation issues to consider.

- Voice-related enterprise applications are identified as those applications that are repeatedly given to or used by employees as opposed to those in wide usage by customers.
- *Hearing versus seeing* output.
- *Directed versus unstructured* dialogue (input).
- *Learning of an application for the first time versus a repeat user.*
- *The underlying "logic" in the systems*. For example, what would be needed to find the cheapest airfare from San Francisco to New York?
- *An inconsistent wireless infrastructure*, producing ambient noise and coverage problems
- A*ppropriate marketing* over pure promotion. User backlash if the technology isn't doing as promised.
- *Spelling has the "e" problem*: e,b,c,d,e,g,v,t
- If *email addresses were numbers*, it would be easier to give voice messages.

The group's collective thinking on economic and business-model issues for the voice applications market.

- *Who pays*?
- U*se existing customer information* for voice database?
- *People want to be able to access email anywhere*. Users must be able to move freely around world, and know that they can manage their calendar.
- *Identify profit models.* Is there money being made per usage of the systems (per call)? Is it a simple request response? Is it simple command and control (horoscope vs. dialogue)?
- *The applications that are likely to be successful are those that drive more transactions*. Technology lends itself to users giving straightforward commands.

- *Reduce costs.* These systems will allow us to replace customer service agents, and increase customer satisfaction.
- *There will be a large market in driving-related transactions.*
- *Telephony costs*
- Greater momentum in the marketplace will create an *economies of scale and the network effect*, so accelerating adoption will make it easier and cheaper to use these systems.
- *There's a need for designers.* Be careful of scarce intellectual resources.
- Twenty-seven percent of people's voices cannot currently be recognized, including children, those with high pitched voices, and some with accents.
- *Capturing a lot of information over the phone is hard to do.* An established context such as shopping replenishment is much easier.

These systems will allow us to replace customer service agents, and increase customer satisfaction.

The group also considered new application areas for voice-related technologies. Common themes are listed below.

Personal Assistant

- Unified Messaging is being improved upon and could be an important application. The "Personal Assistant" would get your email, play it out loud to you, link into your calendar, order by your priorities, and allow you to act upon it There are 18-20 companies in the unified-messaging market.
- Voice dialing directories & directory assistance
- Conferencing by voice
- Virtual private call center
- New phone-related games that leverage outbound calling to draw people in
- Voice portal for personalized access to Internet
- PIM in a voice dialing
- PDA wireless interface (PDA portal)
- Personal schedule and message center
- Directory for call routing (automated attendant)
- Voice dialer, Address book,E-mail reader
- Access to calendar
- Personal response mailboxes
- Content addressable voice exchange
- Time and expense reporting

Customer Service

- Customer care applications
- Customer Relationship Management (CRM)
- Customer service inquiries
- Order status & inquiry
- Appointment scheduling
- Direct voice to data transfer to enterprise DWs, CRMs
- Retail order tracking

Travel Support & Location Services

- In-vehicle telematics services via wireless
- Location-based service for city infrastructure
- "Help me" with directions from cell phone
- Vacation/travel information service
- Location-based voice access to shops and information nearby (temperature and local time)
- 511 (National Travel Info) applications
- Personal location directions
- Airport flight delay and traffic alerts; voice-based messaging
- Severe weather alerts (voice mail, by location)

Sales Force Automation
- Lead tracking
- Field Rep accessing company information/ database portal
- Sales-related information for sales rep
- Sales database applications

Training
- Sales training: voice simulation for responses to objections or for how to create value for certain types of customers
- Enterprise training systems: registration, billing, tracking, certification

Human Resources
- Employee benefits self-service
- HR enrollment & benefits management
- Large enterprise HR systems for handling new hires, transfers, exits

Financial Services
- Bank/ financial transactions
- Brokerage transactions that are voice enabled

Ordering
- Telephone ordering
- Listen to radio from a cell-phone, have it be voice-directed and integrated with voice commerce
- Self-serve for common consumer applications

Field Service
- Field service management
- Field service technician

Corporate Data Access
- Corporate data availability
- Network data access
- Enterprise (private and commercial) voice portals
- Speaker verification to access confidential company information
- Voice access to company intranet

Procurement
- Replenishment to existing customers
- Auction notification
- Industrial inventory
- B-to-C procurement
- Purchasing system for frequent order items (e.g., computers, stationary)

Location-based Asset Management

Translation
- Language translation – due to multiple nationalities living around. One calls and speaks in one's native language. The system translates into local language and repeats to you in your native tongue.

Ideas for Future Think Tank Sessions
The following is a list of future Think Tanks Sessions that the group thought would be of value to the markets.
- Future business models for voice
- Evolution of personal agents
- What is working and where
- Educational needs of schools, and of the marketplace
- Designing compelling voice applications (What makes a good VXML application?)
- Educating users of voice applications
- Defining the value chain of voice delivery. Where is the profit?
- Imagining what happens when telephony costs drop to zero (too cheap to meter, but never too cheap to tax)
- Consumer acceptance of speech technology
- Examining good industry associations. Co-opetition leads to industry momentum.

These proceedings were written by John Willliams of Harbor Research.