CUSTOMER MANAGEMENT insight

THINK SERVICES
A DIVISION OF UNITED BUSINESS MEDIA

- SHOPPING CART
- LOG IN
- MEDIA CENTER
- MAILING LIST
- CONTACT US
- HOME

> EVENTS > TRAINING > CONSULTING > NEWSLETTERS > WEBCASTS > BLOGS

Magazine
New Customer Management Insight
Digital Publication
Call Center Magazine Archives
FREE Webcasts
Podcasts
Vendor Guide
Multimedia Content
Online Advice
eNewsletter Alerts
ICMI Forum
ICMI Jobs
Training
Consulting
Newsletters
Publications/Tools
Industry Research
QUEUETIPS
Conferences
ICMI Global Report
Call Center Knowledge Online
Join ICMI
Members Only
Vendor Guide
Industry Information
About ICMI
Download the ICMI Catalog
_

TechEncyclopedia

Define It

The Portal Pushers: Speech Vendors Poised To Grow The Voice Web

Voice access to websites started as a cute option on UM systems. Two years later, there's a voice mark-up language and a move to hosted voice sites that might actually grow into an industry.

By Ellen Muraskin

12/05/2000, 12:00 AM ET

Remember when the speech technology business looked upon the world wide web with fear, because it threatened IVR? It looked like bad news for text-to-speech (TTS) and automatic speech recognition (ASR) in the telephone network. After all, most of us grasp information more quickly and easily through our eyes than through our ears. And after five to six years of explosive growth in websites and the search engines that index them, the visual web provides more information than any mortal can ever hope to use.

It's yet to be proven, but speech vendors are now claiming that not only will the web not kill off speech technologies in the telephone network, but it will positively ignite them. To quote David Thomson, business development manager at Lucent Speech Solutions, "We're seeing the Internet accelerating the growth of speech rec, by giving us ready access to all the information we need."

Whenever we need it, on any device. "With handheld devices, the visual readout is very nice, but no one is going to want to type on their little pocket PC if they can help it; they'd much rather use voice commands. Even visual web-enabled phones are going to lead to greater demand for speech recognition," says Thomson.

What kinds of information needs are going to be so urgent that they can't wait for the screen in the office or at home? Messaging, obviously. Not just access to voicemail, e-mail, verbal instant messaging, and other media, but a whole slew of

▶ UTILITIES

🖨 print this article
✉ e-mail this article

▶ RELATED LINKS

Why Is It Taking So Long For Speech Technology To Catch On?

Cultivate An On-Demand Workforce Through On-Demand Technology

Executive Interview: Intervoices Ken Goldberg

Nuance Releases Recognizer V9

Avaya Releases Customer Interaction Express

Spring Cleaning: Its Time to Clear the Cobwebs from Your Center

Nuance Unveils Voice Search

Microsoft to Acquire Tellme Networks

Free CallCenter Insider Newsletter

Your Email Address

Optional Areas of Interest

☐ International New s
☐ Advice/Tips
☐ Technology
☐ Agent Development
☐ IVR

commerce- and alert-related messaging and "push IVR," stimulated by the need to keep customers informed about transaction progress or make them aware of opportunities (to act, intercede, buy). Wouldn't it be nice to get an (automated) call from your travel agent, explaining that your flight has been delayed by two hours? Wouldn't it be great if - in answering that call - you were able to book a new flight via voice command? Companies like Par3 (Seattle, WA-206-902-3900) are working hard to make verbal alert messaging and speech-driven push IVR a regular part of the Net's commerce environment, especially the mobile subset of same.

Beyond this, travel-related needs loom large as speech-driven Net killer apps: Where is the nearest taxi? Is my flight on time? Where are hotel rooms available? How do I get from here to there? What's the weather in El Paso? Travel - and the sheer inconvenience of being far from familiar places, equipment, people, and resources, and pressed for time - is also likely to spark additional demand for connectivity solutions (e.g., phone-book lookup, dial-by-spoken-name) and 'on the road' mobile-commerce paradigms (e.g., send a gift basket to the client I just met with) that keep hands free and keep schedules clear of chores.

Stocks are another common application: You may want to check your portfolio, or portions of it, in the midst of a particularly volatile Nasdaq gyration. Sports fans may want frequent updates on scores. All these information requests are natural fits for spoken requests; unnatural fits for DTMF menus.

It's not surprising, then, that the first voice portals - services that offer a range of information sources from one 800 number - are offering a mix of stocks, weather, news, sports, business locators, and audio feeds of headline news and traffic reports. It's also important to look beyond the U.S., where over 40% of homes are currently connected to the web, to Europe and beyond, where PC penetration is lower. There, the phone may be the only medium of web delivery.

Now that delivery can enjoy the amazing distribution model of the Internet, the requested data - and the application itself - no longer needs to reside in the same location as the voice and telephony resources. Companies that want to provide public or password-protected access to their information over phones will not have to invest in telephony infrastructure, any more than they had to buy web servers to launch websites. Instead, voice sites will be hosted: in the network by telecom carriers; by traditional voice service bureaus, to be rechristened Voice Application Service Providers; by ISPs who see the natural multi-media hosting opportunity; and by startups who see themselves as data storage and presentation ASPs.

Here Come The Voicewebmasters

It stands to reason that if voice-enabled web servers and application servers serve up the information, a markup language will be adopted to

enable all kinds of voice-webmasters to voice "publish." That's the promise of VoiceXML, the XML derivative whose 1.0 specification was ratified by the VoiceXML Forum only this March. (See Robert Richardson's excellent look at building VoiceXML applications, "Roll Your Own," in *Computer Telephony* September)

VoiceXML is poised to be to the audible web what HTML was for the visible one. VoiceXML pages can now be uploaded to web server test beds and hosts, just as HTML pages are today.

Result: Nine months after VoiceXML 1.0, it is now raining VoiceXML interpreter products.

VoiceXML

While the VoiceXML spec itself is quite limited in scope, an integrated VoiceXML browsing platform, such as now being assembled by Lucent and other vendors, does take care of a lot of the telephony housekeeping that has thus far discouraged potential developers. Dave Thomson, in his VoiceXML 101 presentation to Internet Telecom Expo attendees this September, started with a three-line program that did "nothing" but answer the phone and hang up. But as he pointed out, that "nothing" program:

```
>
>
>
>
```

is all his VoiceXML-interpreting speech server needs to see to answer a call, look up a URL to match a DNIS, fetch a VoiceXML page, execute a VoiceXML script, and hang up. Software added by Lucent performs the DNIS-to-URL table lookup and accesses the page.

VoiceXML is so easy, says Thomson, that we can look forward to the end of the terrible drought of CT application designers, because armies of webmasters will take up the new tool. Developers will simply design their VoiceXML apps, perhaps test them locally using PC mikes, and FTP them up to networked voice resource DSPs and cards living elsewhere on the Internet. Then they'll test them further, and "publish" them to the world.

If it truly follows in HTML's footsteps (and even desktop publishing's, ten years earlier), we'll see a lot of amateurish human interface design. "VoiceXML is going to make speech so easy that it's going to be very easy to go out and shoot yourself in the foot," concedes Thomson. "People aren't going to consider what you need to do to make a speech recognizer work properly: constrain the vocabulary, set the timers properly, make prompts unambiguous. Web page designers aren't necessarily going to have any familiarity with that."

"But that's a good thing," Thomson hurries to add, "because if you have lots of people going out and shooting themselves in the foot, eventually they're going to learn to aim. And it's going to become a huge business." Then, presumably, we'll see standards rise to the level of the best speech-

enabled IVRs that have been programmed in more traditional ways, through Javascripts, app gens, and C++ code.

Definitions

Like all new buzz, he Voice web suffers from a great blurriness of definitions, which hopefully will standardize along with their specifications. Here's our attempt: **Voice portals** are services that give subscribers access though one PSTN number to a variety of different content sources. Text-to-speech is an essential component of that service, and of the web-serving platform that accesses URL pages or http-delivered data and reads out variable fields. Speech recognition would appear to be essential as well, to navigate between many content sources and to provide the user input. At least one voice portal service now in existence, however - Yahoo-by-Phone - is still using touch-tone input at presstime.

Voice portals also refer to the platforms being offered, sometimes incorporating the speech technologies, sometimes leaving them on a separate server - which read and execute the speech-enabled IVRs that constitute the services. Portals may host the service locally, or fetch pages or individual data fields via http.

Defining **Voice browser** is trickier, since it isn't neatly analogous to Netscape or Internet Explorer. The phone set is an extremely thin "client," more analogous to the user's keyboard and mouse than a whole HTML browser. With voice, the true browser lives on the server side. The voice browser is a server-side application that conducts the user interface to the portal: it understands commands like "back, forward," it listens for hot words that will yank the speaker back to the main menu of services, and performs navigation to sites via URL or call transfer. A voice browser, just like Netscape or IE, should also hold user preferences - favorite sites, stocks/ weather reports/sports teams of concern, to be used the way a visual site might use cookies. The browser may perform speech verification to ensure identity; alongside this comes walleting services; holding the verified speaker's credit card information to authorize mobile purchases.

To most, the voice browser is also the part of the platform that does the VoiceXML interpretation; that renders the VoiceXML tags into lower-level languages (Java), which in turn directs the telephony and speech drivers. This makes the voice browser a software piece with which to upgrade preexisting IVR platforms. A better term for this is **VoiceXML interpreter**.

So where do voice portals stand as of October 2000? Announcements of new entrants arrive daily. Of those that provide access to multiple information sources, many of us have heard of **BeVocal** (See March "First Looks," and try 1-800-4BVOCAL), which officially launched this summer with a cleverly rendered, custom programmed set of services to sports, stocks, weather, travel directions, and lately, business locators. It remains independently owned, but has begun to host services for carriers. It uses Nuance ASR.

Tellme (1-800-555-TELL) launched this spring and provides a similar selection, plus Tellme studio (see Robert's article, again), a testbed for

designing VoiceXML applications. **Mobilee** has signed a contract to voice-host traditional-web portal Lycos, and is operational in five cities as of this writing. Both Tellme and Mobilee are Nuance customers.

In the Speechworks camp is **HeyAnita** (1-800-44-ANITA), which officially launched October 3. It offers news, flight information, weather, sports, stock quotes, horoscopes, and a personalized service called Message Center, which reads e-mail. More services, such as V-greetings and wake-up and reminder calls, are planned for 2001. There's also Quack, which was purchased by AOL before it launched; it is now hitting the market as **AOL by Phone**.

In Italy, Philips Speech Processing has bragging rights to **Omnitel**'s portal.

Info by Voice, from BellSouth for Atlanta, has been around for a while and predates the VoiceXML boom. It offers stock quotes, NBA, NFL and college sports scores, weather, a restaurant guide, horoscopes, soap opera updates, and other information. It uses SpeechWorks natural-language recognition.

Nuance cites the Kelsey Group's forecast that the U.S. market for voice portal services and technologies will exceed $12 billion by 2005. They cite U.K.-based Ovum predicting a world market of $26 billion. Allied Business Intelligence finds 56 million mobile voice portal users in North a by year-end 2005; 250,000 voice sites, and a $50 billion v-commerce market!

But not only is the jury still out on the success of the voice web: the jury hasn't even been assembled. Most people outside the industry still haven't heard of voice portals. Nuance's Chuck Bullam, addressing a Natural Microsystems' Alliance Partner Conference in Boston in October, described voice portal pioneers such as BeVocal and Tellme as "flourishing." But pressed to define "flourishing," he can only speak in terms of venture capital accrued and partners signing on to be speech-enabled. Business models are in the experimental stages, analysts and vendors agree. And most all analysts agree that fifty different 800-number portals, all giving access to the same stocks/sports/weather/traffic services, won't all survive.

The greater likelihood, say several analysts, is that voice portals will find niches if they're acquired, and if they target specific localities or user groups - teenage girls or sports fans, suggests Gartner Group's Jackie Fenn. Another possibility is the use of voice portals to complement standard web portals. This can already be seen in Lycos' contract with Mobilee, and in AOL's acquisition of Quack.com. Here there's a prexisting user base, many of whom have already indicated their information preferences over the web.

The AOL Of Voice

BeVocal (Santa Clara, CA-408-748-8700) started out aiming to be, in founder Amal Joshi own words, "The AOL of Voice." But guess what: AOL decided to be the AOL of voice (1-800-AOL-1234). It then took the shortest route to that path by buying itself Quack.com. BeVocal has

gotten the message, and now sees its greatest opportunity in rebranding or co-branding its service for existing web portals, carriers, and car makers a la OnStar portals.

We can only assume that it's been harder than BeVocal expected to garner subscribers or sponsors for an advertising-based model; instead, they'll take the technology to the folks who already have subscribers. Joshi says that deals are imminent with two major North American carriers. The first became public on Oct. 30,: when Qwest Communications announced a voice portal service, hosted by BeVocal, for its wireless subcribers. (See News, this issue).

What's the lure to carriers? A chance to lock in customers with personalized portal services, a la MyYahoo; an increase in wireless minutes of use, subscription fees for bundled-in upgrades such as voice-activated e-mail, stock alerts, voice dialing, and perhaps content-specific sponsorships; a personalized stock reader that's prefaced with a word or two from Schwab, for example.

Joshi sees the host-carrier relationship evolving, from one of quickly outsourced v-presence to eventual buildup of in-house ports. When that happens, BeVocal will be happy to license their SpeechObject technology and extensive work in grammars and user interface design. To take just one example: Do you want to build your own grammar and dialogs to recognize 45,000 U.S. cities, or would you rather get the Java Bean Speech Object from BeVocal?

Joshi predicts that the first wave of voice portal applications will be access to content. That will be followed by v-commerce, ticketing, and enterprise applications like voice-dialing and Extranet access. In the meantime, BeVocal is hosting its own third-party developer program, offering their VoiceXML validation tools, SpeechObjects that can be plugged into VoiceXML pages, and their own testing environment.

Carriers: Lend Me Your Ears

Telcos not only have the advantage in already having subscribers: They also already have the advertisers, through their Yellow Pages and now website publishing arms. So the data is waiting, if not yet XML-ified, for regional talking Yellow Pages to be provisioned through the portal. BeVocal has a great proof-of-concept here (and outsourcing offering) in its recently-launched BizFinder service, which finds the nearest location of 2,000 brands now, in 1,000,000 locations, according to Joshi.

The business model here is not only advertising/sponsorship, but slotting, or placement fees; portal services can charge premiums to restaurants, say, who come up first, not third or fourth, in searches for a particular cuisine or location.

Voice Portals As Churn Burners

The carriers' great hope for portals is as churn preventers. And the best way to bind the subscriber to the portal, as with myYahoo, is personalization. Just as Yahoo, Excite, and Netscape.com invite users to adopt their sites as browser home pages with configured stock portfolios,

sports teams, and weather interests, voice portals can use HTML interfaces to learn what their subscribers want to hear about. When voice-dialing is added to the service, it's also easy to use the traditional web to upload personal address books, either from Outlook or other clients or even, via serial link, from PDAs and Palms. This has been possible ever since phonetic speech recognition (using dynamic, text-entered vocabularies) arrived, about three years ago.

"Every conference on wireless talks about increasing average revenue per user," says David Bradshaw, of market analyst shop Ovum. "Mobile companies have paid enormous sums for spectrum and are all finding diminishing customer revenues. Part of their plan, long-term, is to become the subscriber's 'trusted confidante,'" he says. This entails enabling v-commerce by storing the subscriber's credit card number, or being the conduit to all sorts of personalized content sites.

At Italian mobile operator Omnitel, one of Philips Speech Processing's star users, this has worked out with a portal that offers over 70 services. "People are using their phones a lot more," Bradshaw reports. (Although Omnitel users, when polled beforehand, said they'd make the most use of other services, the far-and-away favorite service for Omnitel is first-thing-in-the-morning horoscopes.)

Service bureaus, too, may be happy to take on portals if they can avoid the freight charge by pulling in VoiceXML pages over IP, instead of transferring the call to PSTN-based services, as is now the case for some portals.

It's important to remember that VoiceXML isn't only for portals. VoiceXML also looks promising for applications that will sound to users like traditional, speech-enabled IVR: 800 numbers to one content source. Callers won't know the difference, but their calls may be powered by Voice ASPs VoiceXML interpreter/web servers, fetching VoiceXML pages that have either been uploaded by developer tenants a la HTML, or dynamically fetched across an IP network.

"Closed Garden"

Voice portals that roam the web freely - in the sense that your Netscape browser can surf ceaselessly from .com to .org to .fr, .de, .nz and so on - do not exist at this point. To date, all the portals are closed systems, giving access to only those content providers with whom the portal providers have agreements (and IP- or PSTN-based routing tables). All current portals also attach to only one "browser." That is, to one set of telephony and speech resources that terminate the PSTN call; these platforms either execute regular IVR applications or transfer and conference in other destinations.

Speech companies - those in core technology, app gens, hardware, software, and combinations thereof - are now drawing up a range of scenarios for growing a truly surfable voice web.

Speech Technology Vendors

Nuance

Nuance's (Menlo Park, CA - 888-nuance-8) V-Builder, announced and demoed to great fanfare in October 99 (See "Nuance's Voice Browser," *Computer Telephony*, November 1999), has just reached general availability, says product manager Brian Kaplan. It allows developers to look at an existing HTML site and, interaction by interaction, choose the best Speech Objects to voicify the application and render it in a combination of VoiceXML and Speech Objects. Apps can also be tested on the desktop with this tool and a PC microphone. V-Builder comes with 25 reusable Foundation Speech Objects, and over 100 pre-packaged grammars. V-Builder's VoiceXML product can also be amended with standard Java editors.

Nuance was also first to market with a productized voice browser, Voyager. Megan Gurley, Yankee Group analyst, goes so far as to say, "Voyager is what started the voice portal craze."

Nuance's Developer Network, with its $495 membership fee waived in late September, has served over 3500 developers evaluation copies of their SDK for the web, through Nuance's Extranet. The SDK includes Nuance's foundation Speech Objects, which are their containerized speech-recognition Java Beans for common ASR tasks: dates, currency amounts, social security numbers, and the like. According to Kaplan, new members are joining the network at rate of 20 per day.

Nuance has also approached ASPs, carriers, and ISPs with a Voice Web ASP program, an offering of products, services, and training to help ASPS host and deliver voice-enabled apps of all kinds. Part of the carrot here is bundled pricing for software licenses, business development and lead referrals, training, tech support, and co-marketing programs. They're also encouraging members to serve as Voice Site Staging Centers - Voyager-equipped sandboxes for developers producing VoiceXML and Speech-Object-based applications.

The Voice Web ASP program offers twelve deployment ports of Nuance 7.0 core recognizer, Verifier 2.0 for voice authentication, the Voyager browser, and Nuance Order management Suite for V-commerce. Also: six Nuance Developer Network memberships, and V-Builder., the VoiceXML tool.

And finally, Nuance is also working to grow the market for potential Voice Web tenants, by inviting them into the initiative through the "Voice Web Content Program." The offerings here are very similar to the ASPs, for those companies who wish to turn their webmasters into VoiceXML- and Speech-Objects-fluent voice webmasters. It includes Nuance consulting services. It's not only to encourage companies to bring up voice sites, but to offer syndicated content for dissemination by multiple portals. Charter members include AccuWeather, Astrology.com, Audible.com, Blue Wireless, BuzzHits, CNET News. com, Etak, HOLLYWOOD.COM, Indicast, InfoUSA.com, iSyndicate, IT Network, Menus.com, ON24, Televigation, Voquette.com, and Zagat Survey.

SpeechWorks

"The voice web so far has an 'open problem,' I can only convert to those

services that are known to the browser. We need browser-independent connective tissue - one level above VoiceXML - to jump from one speech service bureau to another," says Stuart Patterson, CEO of ASR (and now verification and TTS) provider [SpeechWorks](#) (Boston, MA - 617-428-4444). "Tromboning" keeping one termination open and then bridging in a second leg to another site is not the only model of the voice web," he says.

Their answer, provided as open source, is [Speech Links](#); a SIP- and XML-based API to be adopted by all voice sites whose voicemasters, let's call them, want them to be linkable. Speech Links will some day, the thinking goes, send the call itself and data about the call to another site, sitting on another VoiceXML interpreter.

The spec for SpeechLinks is a pdf you can download from SpeechWorks' site. The VoIP protocol here is SIP; so it'll take a while, no matter how fast SIP is adopted, till SIP user agents proliferate at destination voice sites. But that's the thrust of this initiative, and SpeechWorks says that code insertion to link up preexisting sites should not be a big deal.

For the present, the spec only has the voice site's SIP user agent returning a session ID and a phone number in the body of the SIP acknowledgement. The voice server then uses that phone number to execute a second-leg PSTN call to the next browsing destination.

Peer deeper into the medium-agnostic nature of SIP, and SpeechWorks comes up with the mother of all screen-pop scenarios: A caller Speech Links to a travel site, decides to speak to a real travel agent, agrees to transfer the call (with caller profile information) there, and her caller profile info shows up in the agent's HTML browser at the same time. A full SIP implementation of both voice- and data-over IP would make it possible to release the call from the first VoiceXML interpreter and spare the Voice ASP the cost of two legs. This could also hold cookies, preferences, browser history, and navigation pointers. (This has occurred to BeVocal, running Nuance ASR, as well; their presentations also posit multi-modal sessions made possible via SIP.)

In the words of SpeechWork's own FAQs on Speech Links, what makes Speech Links different from the tag in VoiceXML is that it specifies the telephony and the IP handshake necessary to accomplish a transfer. The Speech Links provide a unifying mechanism that can establish a telephony (phone number) and data (URL) link with call control and the ability to correlate a data stream and voice call.

Speech Links are also designed to let any VoiceXML interpreter support call transfer or receive between any service, VoiceXML-based or not. In the latter case, it would activate traditional PSTN call transfer.

On the browser beat, SpeechWorks' answer to Nuance's Voyager platform was its Open Speech Web initiative, announced at the end of June. SpeechWorks is now going to offer their browser/ VoiceXML interpreter application for free, on an open-source basis, following, they say, the Mozilla model of the visual browser. They position it as an effort to ensure standardization. "We want to avoid the Balkanization of

browsers and VoiceXML," says Patterson. The browser will still need an underlying ASR (as well as TTS) engine to run, of course, as well as telephony hardware. (Qualifying developers may find some or all of these pieces in free evaluation voice portal platforms being distributed jointly by Dialogic, SpeechWorks, and Nuance.)

The SpeechWorks Speech Browser includes the VoiceXML interpreter, with Java and HTTP elements to support linking to websites, and handle grammars and prompts as defined by the VoiceXML standard. It also comes with a SNAPS Layer - speech navigation and presentation scripts, which include "hot word" customization and event log history. This means that the interpreter continually listens for and responds to a "hot word" that can grab the caller back to a main menu from a specific site. The browser also comes with base files so that developers can quickly link specific prompts and vocabulary to specific URLs.

Recognizing the importance of establishing an open test bed, SpeechWorks has given the VoiceXML platform first to Carnegie Mellon University to host. Both Speech Browser and Speech Link open source code are to be made available from Carnegie Mellon's download site in Q4 2000. The Speech Links spec is now available at speechworks.com.

SpeechWorks also wanted us to know that they have licensed AT&T's large-vocabulary text-to-speech; that AT&T's extremely sturdy TTS has been massaged, fine-tuned, scaled, documented, and branded by SpeechWorks into a product called Speechify. SpeechWorks recognizer itself is up to Version 6.5 since August. It boasts new Dialog Modules, including speaker verification, address collection, and voice enrollment capabilities. Version 6.5 also introduced new Natural Language tools for editing grammars, vocabularies, and required pronunciations.

Also: First Union Corp agreed to deploy 3,700 ports of SpeechWorks' ASR for an account query and bank transaction application. A top Taiwanese brokerage company, MasterLink Securities, has launched the first Mandarin-speaking stock trading ASR application, and VoIP powerhouse Net2Phone and AT&T have become investors in the company, which had a successful IPO in August.

Philips Speech Processing
Philips (Dallas, TX - 972-726-1200) supports the largest world-wide deployment of a voice portal in Italy's Omnitel mobile carrier. It also announced the world's first Mandarin-speaking voice portal with KG Telecom in Taipei: 18 services, including financial and traffic updates, taxi services, airline reservations, pizza orders, and movie reservations. Philips is also behind Cgtel, the number-two mobile provider in France, aggregating 20 services.

In the U.S., however, Philips is not the strongest part of the Voice Portal Hallelujah chorus at this point. "In the Americas, we're waiting to see which voice portals and economic proposals make sense. We don't have the pressure of some tech companies, who are holding enormous market caps, so we're freer to go after less visible applications, where we think there's a more economic proposition," says Pete Foster, Philips Speech Processing's executive vice president. "The voice portals are a much

higher risk than directory-assistance call center or voice-enabled dialing," where Philips has had a historical lead, and where twelve to 14 people are dedicated just to adding names for directory-assistance buildout. "We're waiting to see which voice portals make sense."

"The direction we've taken is to go after the more pragmatic applications. We're seeing thousands of ports going into improving mobility services like plain old voice mail. We are now in some 15 different carriers through a number of partners, the largest of which is Comverse, doing voice-controlled voicemail. We're seeing this really picking up speed in deployment," say Foster. Then there's call-center offload: "I was just at a CLEC's with 118 exchanges; their problem wasn't finding customers; it was finding people to service customers." The full employment picture is forcing call center employers to look at self-service, speech-enabled IVR alternatives more tolerantly.

The Omnitel experience has taught Philips that "There are certain things people will pay for," says Foster. Although the portal offers over 70 services, according to Foster, 85 percent of traffic goes to eight, which include the aforementioned horoscope, lottery results, and sports scores, especially during major events.

It's a very sophisticated portal, says Foster, with multiple levels of billing. It sells rechargeable prepaid cards to get to premium services. You can also buy movie tickets and make airline reservations. Omnitel gets a commission on sales and as noted, has seen a significant uptick in minutes.

To recap, Philips' core products start with Speech Pearl, the core recognizer, now good for 200,000-word flat-file vocabularies. It now runs on Periphonics/Nortel's Oscar voice peripheral; one deployment of this is in BT's Directory Assistance and call completion. SpeechPearl last July added Greek, for a total of 20 languages. SpeechPearl also can switch between different languages instantaneously, within the same application. The whole development environment-plus-core-recognizer is Speech Mania; this includes recording tools, vocabulary building and tuning, and writes all the way down to the APIs of telephony hardware.

Philips' large-vocabulary products are also known for their Natural Language processing, which allows an application to derive four or five variable answers from an unstructured caller request ("I'd like to fly to Barcelona on United next Monday at three PM,"), figure out which piece might be missing, and ask for that ("From what airport?").

For smaller vocabularies, there's SpeechWave, which is DSP-based and is used for more narrowly focused apps such as credit card or calling card number number rec, voicemail navigation, and speaker verification. SpeechWave comes in more than 35 languages.

Locus Dialogue
Speech recognition company Locus Dialogue (Montreal, QC, Canada - 514-954-3804) is also offering a voice browser application as part of its Speech Portal hardware-and-software offering. Speech Portal consists of the core Locus Dialogue speech technology, hardware, and software required when integrating Internet infrastructures, packaged speech applications, and telephony infrastructures.

Locus's booth at Internet Telecom Expo featured partner SpeechWise Technologies, (San Francisco - 415-409-4400), which was talking about its Voice Applications Platform (VAP) for faster deployment of voice-enabled web applications. SpeechWise's VAP incorporates reusable components, a reference architecture, and an application development framework that talks to Locus' platform on the telephony side; on the data side, it talks to Apache, IIS, and NES web servers and incorporates existing business logic. Runs on Linux, Solaris, or Windows NT, in a J2EE-compliant JSP and servlet environment.

SpeechWise was also offering two prepackaged applications to run on Locus' browsing Speech Portal and SpeechWise's VAP: Speechwise Auction, to voice-enable auction sites like Ebay, taking bids, checking bid status, and listing all active bids; and Speechwise Calendar, for scheduling and notification.

Server Solutions

IBM Voice Systems
IBM Voice Systems (West Palm Beach, FL - 914-642-3000), like Comverse Network Systems, has bolted a VoiceXML interpreter and its own ViaVoice technologies onto its preexisting DirectTalk IVR platform. It also makes a good marriage with IBM's Websphere.

IBM's Websphere is a web server (based upon Apache) plus an application server; it therefore serves up dynamically assembled HTML pages. It sits in the middle layer between database and the presentation layer, which traditionally is a Netscape or Microsoft IE web browser. It executes Java servlets and runs Enterprise Java Beans. IBM quotes a March IDC report that gives WebSphere 9.5 percent of the application server software platform market at the end of 1999, up from .4 percent the year before. Giga Group, in a forecast published in June, said IBM will have 24 percent of the Java-based application server market by the end of 2000.

IBM is now planning to launch a WebSphere Voice Server (WVS), built on DirectTalk, to put another presentation option in front of the application-and- page-serving WebSphere. It's scheduled, at press time, to launch at Voice Europe in the beginning of November. It will thus deliver VoiceXML-interpreted content over the phone, as well as IVR programmed with other tools. WVS gets VoiceXML pages via http from either the Websphere Application Server (WAS) or anyone else's application server.

On the DirectTalk platform, DirectTalk Beans for Java give the developer a way to develop ASR and TTS-enabled applications. Atop that they've put VoiceXML, to direct the speech drivers another way. You also have use of DirectTalk's State Tables; a very good discussion of the relative merits of the three application programming approaches is given in their DirectTalk pdf, available soon, I am assured, at www.ibm.com/speech. DirectTalk comes on AIX, and is available for U.S. and U.K. English, French, and German.

The WebSphere Voice Server also runs with the Cisco Voice-enabled

router, which speaks JTAPI to its H.323 stack. It is thus the first VoiceXML interpreter to run with a VoIP client, they say. The core recognition engine and TTS live in the WVS, and respond to JSAPI calls being sent by the VoiceXML interpreter. The WVS for Cisco runs on NT.

Says Sunil Soares, program director of product management, IBM Voice Systems: "We needed to move DirectTalk to the web world - so we speech-enabled DirectTalk. We added speech recognition and text-to-speech, and we added DirectTalk Beans for Java as a free download for developers. Then we pushed the envelope further. We decided to offer VoiceXML, so that developers can use state tables, Java, or VoiceXML to build applications."

AGENCY.COM is a high-tech marketing company that helps clients build and grow multimedia interactive businesses across multiple channels - the web, wireless, and interactive television. The company is an early referral site for the IBM WebSphere Voice Server, and has found it effective in linking non-traditional devices to web-based applications.

Again, IBM Voice Systems has traditional customer design wins that speak to the strength of their core recognition technology: Soares told me about Cable & Wireless, which was using their Via Voice directory dialer for 14,000 names in an enterprise. Brokerage giant T.Rowe Price is also using DirectTalk IVR for 300 to 400 thousand calls per month for trading, using American English ASR. And in Nova Scotia, Scotia Bank is piloting speech rec on top of DirectTalk to handle 700,000 calls per week. Ninety percent of these transactions are handled completely by the TeleScotia IVR, speaking English, French, and the Chinese dialects of Cantonese and Mandarin.

Lucent Speech Solutions
Lucent Speech Solutions (Murray hill,IL - 630-979-2000) has now worked the VoiceXML 1.0 spec into its Speech Server, and is very actively promoting both the standard and the server to carriers and ISPs. David Thomson, CTO at Lucent Speech Solutions, gave an excellent VoiceXML tutorial at CMP Media's Internet Telecom Expo in New York this fall. Dan Krasinski, business development manager, was seen at NMS's Alliance Partner Conference in Boston in October (along with many other shows), presenting Lucent's CO-grade Speech Server in its VoiceXML-interpreter role.

By virtue of its AG4000C NMS cards for T-1 interface and custom speech board, Lucent's speech server is high-density; up to 192 channels, or eight T-1s. It's got the 100M bit LAN interface on the data side, and, of course, Lucent's own TTS, an engine we at *CT* Magazine rated highest last year for intelligibility. It also uses Lucent's ASR and speaker verification, and supports TDD as well. It's NEBS-compliant and CompactPCI.

The speech boards are powered by three Power PC 750s; up to eight of these can go in the 19-inch Radisys CPCI CP80-15 chassis that Lucent uses to package up the server. The host CPUs are dual 600 MHz Pentium IIIs. They can be delivered, stacked for traditional DC telco use,

in a 24-inch 5ESS cabinet, as well. A typical server, says Krasinki, houses two NMS AG4000Cs and eight Lucent speech processing boards. Available on Sun Solaris.

Authoring tools for the Speech Server include, at the highest level, VoiceXML and XML and the Java Speech Grammar Format (see figure. Developers also have use of Java Finite State Machines, as well as Java RMI for remote applications and commercial packages such as JavaMail and JDBC. Deepest down, purists have C++.

Lucent has placed a speech server outside their firewall to serve as a test bed for VoiceXML experimenters. Lucent has also built an extensive web-based monitoring and logging option, which runs on a separate server. It gives real-time health-status from the server cluster level down to individual servers, telephony interface cards, speech cards, and individual ports. The OA&M server displays alarm summary views and offers descriptions of failures, and also illustrates dependencies; as, for example, when a VoiceXML interpreter depends on the health of a web server or e-mail server.

Comverse Network Systems
Comverse Network Systems (Wakefield, MA - 781-246-9000) has announced a portal platform called Tel@Go that incorporates voice-activated dialing, web-based profile entry, e-mail reading, web browsing to user-selected sources, and other information. CNS has found its first portal customers from among its existing 340-network operator customers, who already use their enhanced services platform.

CNS's integration effort, of course, varies with the preexisting installation. "If you have a Converse unified messaging system that already has message aggregation, TTS, a web interface, and an address book, we expand what is in the address book to add voice dialing, personal preferences, and add a module for natural language understanding. That unit runs the Tel@go application and speech recognition and has ports to answer calls," says Michael Krasner, general manager of the speech portal division at Comverse Network Systems.

It's a very carrier-centric kind of portal, with an emphasis on voice-activated dialing and message retrieval. A typical user will dial in, hear messages, dispense with mail, and then, via TTS, hear the sports, news, weather, etc. from the sources that he's specified on his web form.

"VoiceMarks" are Tel@go's equivalent of bookmarks. Krasner says that information-centric deployments are also part of their offering. To date, the portal is available in various English dialects, French, German, Brazilian Portuguese, Italian, and Spanish.

The portal platform has announced deployments with Nuance and Philips speech recognition, but will work with other recognition engines, too, to suit operators' preexisting choices or other considerations. VoiceXML compliance is coming, but not here yet, says Krasner. They have a huge customer win in Sprint PCS, with access to a selection of content sources. Krasner reports that more carriers are now asking the portal platform providers to arrange content sources as well. Toward that end, they are working with content suppliers and aggregators.

Genie Technologies
[VoiceGenie Technologies](#) (Toronto, ON, Canada - 416-736-0905) announced their VoiceXML platform at CTIA Wireless on Oct. 16. They also have a developer network - register at developer. voicegenie.com - where tools, a testing platform, documentation, and developer forums are posted.

Board Vendors Package Up Portals, Too

CT board makers Dialogic/Intel and Natural Microsystems have also recently announced voice portal platforms.

Intel/Dialogic
[Dialogic's](#) (Parsippany, NJ - 973-993-3000) Voice Portal Platform is a bundled offering with a finely targeted market. It's selling briskly (as in hundreds of smaller-sized development systems) to major ISPs, says Rick Renna, Dialogic sales engineer. "It's for ASPs who are building out a data store and want to rack these up to bring the data out to voice," he says.

It comes with two Pentium III processors, an Intel 440GX chipset, two PCI slots, four hot-swappable hard drive bays, an integrated 10/100 network controller, and analog and digital Dialogic voice resources in the form of the (new-since-July) JCT-series voice-plus-network interface cards.

These new cards, with Dialogic system Release 5.1, will come with soft fax as well, and incorporate the Continuous Speech Processing (CSP) technology announced last May. CSP consists of algorithms that offload high-quality echo cancellation, voice activity detection (VAD), and pre-speech buffering from the host processor, screening data as it arrives and alerting the host system only when it detects voice traffic. It allows the host to process more concurrent speech ports.

The CSP algorithms are built in spec with SpeechWorks and Nuance recognizers. Systems now being shipped are leaving the choice of ASR up to the developer; TTS is L&H's Realspeak. Developer kits are going out with 36 ports; one T-1 network interface and twelve analog ports. Production units will have dual-span and quad-span JCT-series boards, for up to 96 ports on T-1s or 60 ports of E-1 ISDN.

Developer kits are selling for $8900 for the T-1 with analog, $9900 for E-1 with analog. They are non-discountable, and available through distributors as well as Dialogic. Deployment platforms range in price from 12K to 25K, depending on volume, according to Mike Ehrlich, Dialogic's speech product manager.

Dialogic Voice Portal Platforms currently leave application authoring environment and VoiceXML interpretation, as well as ASR engine, up to third parties. Dialogic does have a partnership program in place, however, with those seriously interested in deploying voice portal platforms with Nuance or SpeechWorks: Potential customers may, at the companies' discretion, have free use of the developer-kit platform for 120 days, after which they may purchase the platform at discount or return it. The evaluation program is "evolving," says Ehrlich. Expect to see future

evaluation agreements with the speech rec vendors and with application environments, such as Intel/Dialogic's own Parity Software.

Natural MicroSystems
Natural MicroSystems' (Framingham, MA - 508-620-9300) own platform is called Hearsay. It comes with Nuance speech rec (they say they'll add other ASR vendors in the future) and voice authentication; and integrates an NMS PCI AG2000 or 4000 or CompactPCI AG4000C single-slot voice board, for a maximum 120 ports of speech on E-1s. Its uses an API that integrates telephony and speech functions, supports barge-in and on-board echo cancellation. Universal ports let developers expand voice portal apps to include fax, conferencing, and VoIP. Runs NT or Solaris on Intel and SPARC.

Voice ASPs

These companies have assembled speech technology and network hardware, and are ready to host speech-enabled applications and whole portals. Where once they typically did the design work, the future holds the promise of a more hands-off web hosting service.

General Magic (Sunnyvale, CA - 408-774-4200), known for years as provider of the Portico personal assistant service, was seen at Internet Telecom Expo promoting a Voice ASP service on their pure VoiceXML 1.0 MagicTalk platform. They were offering Nuance speech rec and Fonix TTS, but were planning to add IBM's ViaVoice or SpeechWorks for ASR per customer preference, and Lernout & Hauspie (Burlington, MA - 781-203-5000) for TTS.

They told me their platform was a speech- and VoiceXML-enabled IBM WebSphere Voice Server, built atop the IBM DirectTalk Voice system. They also told me that they were the platform behind GM's OnStar Virtual Advisor, the sports/news/e-mail, stock-quoting portal activated through the OnStar button built into GM cars.

General Magic was offering to design the natural language interface for customers. "We're creating multiple personalities and branding for companies," explained Linda Hayes, VP marketing. Voice tenants could also use an off-the-shelf "Mary" personality, that same savvy female doing the prompting and call screening and calendaring for Portico.

Price Interactive (Reston, VA - 703-620-4700) is already the "home" of everyone's favorite voice-enabled web demo, mapquest.com. They're also old hands in the IVR service bureau business, and SpeechWorks partners. PI is now clearly prospecting for its SpeechPort service among the web's major traditional tenants, while building out its hardware and human-resource infrastructure. L&H's RealSpeak is their text-to-speech engine.

In August, the company bought FusionTec, a Virginia-based e-business consulting firm. The company's strength is in implementing Java-based m-commerce solutions. In September, they cornered us at Internet Telecom Expo to shout about their 25,000-port buildout, to support over 35 million minutes per day. SpeechPort carrier-grade facilities in St. Louis, MS, Fairfax, VA, and (future) Denver, CO, feature Sun Netra/Apache