

Speaking Tour: ASR And TTS Follow The Money

Free CallCenter Insider Newsletter

- [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](#)

A survey of the current speech market shows voice-activated dialing and call center offload are still the easiest sell. These may open the door for voice portals.

By [Elen Muraskin](#)

06/05/2001, 12:08 PM ET

Our periodic speech-recognition story has wandered far afield of the basic technology in past four years. Not because there isn't news there. But it's familiar news: We're used to hearing about improvements in recognition accuracy, higher processor speeds, larger vocabularies, more language coverage, greater affordability of speech-enabled IVR.

The new speech news veers away from technology and toward (equally intricate) business developments. The news concerns the increasingly complicated web of partnerships, forward and backward moves taking place among the myriad players in the speech game, as promise and potential slam up against the demands of profitability.

As those who've been following speech technology can attest (and for those who haven't, see our archived stories at [commweb.com](#)), we've stacked up a big pile of promises here. Speech companies - their stock prices in Nasdaq shock and their own earnings still expressed in red - well understand the need to demonstrate ROI to potential buyers. You can sense the sea-change in the new dollars-and-cents emphasis running through PowerPoints and interviews.

Certainly - or so most claim - any speech company that can demonstrate a solid ROI will find buyers. The first article of faith is that "there's money out there." Often quoted in today's speech market is a forecast, from Mark Plakias of the Kelsey Group, predicting speech-related services that reach \$41 billion worldwide by 2005 - \$27 billion in consumer-focussed applications, \$14

► UTILITIES

-  [print this article](#)
-  [e-mail this article](#)

► RELATED LINKS

- [Nuance Releases Recognizer V9](#)
- [Nuance Unveils Voice Search](#)
- [Microsoft to Acquire Tellme Networks](#)
- [Aspect and Microsoft Create Speech Application](#)
- [Nuance Acquires BeVocal](#)
- [Aspect's Voice Portal Receives Nuance Certification](#)
- [Sarbanes-Oxley vs. Hosting?](#)
- [Speech Recognition Comes of Age](#)
- [Aspect Intros VoiceXML Portal](#)
- [Paul English, IVR Consultant?](#)

Your Email Address

Optional Areas of Interest

- International News
- Advice/Tips
- Technology
- Agent Development
- IVR

TechEncyclopedia

Define It

billion in enterprise, and comprising 4% of total telecom spending. Impressive figures.



So for focal point if nothing else, the question to ask this time around is, What's going to make that money in speech? What speech-enabled applications are people paying for? What platforms are selling? To whom? And what about the three biggest, inter-related promises of Speech 2000: the Voice-delivered web, the Voice ASP, and the VoiceXML markup language?

Stand-Alone Portals Are Mortal

Last fall, we were just beginning to see stand-alone voice "portals" - 800-number services offering access to a wide range of speech-enabled "sites" - start adjusting their business plans. Tellme and BeVocal, first to the scene, had realized that consumers would not rush to pay for speech access to large collections of weather, sports, stock, traffic, etc., information. Nor would consumers tolerate the audio advertising needed to make the service profitable.

The answer, these portal services now think, is to partner with companies who already have subscribers - wireless and wireline carriers, large ISPs, etc. - and offer their platforms several ways: as carrier-branded voice portals, or as speech-enabled IVR hosts. In due course, we've seen portal/host announcements: BeVocal supplies branded voice browsing and voice dialing for Qwest Wireless; Tellme will do likewise for AT&T Wireless; [HeyAnita](#) (Los Angeles, CA - 323-692-1555) will do Sprint PCS. And web content providers are taking up similar tenant roles: AOL-by-Phone is brought to you by what was Quack, another portal.

The hopeful news is that these carriers and content providers, after months of offering voice browsing to subscribers on a free-trial basis, are just now beginning to charge for it (see individual writeups, below). But they haven't dared charge just for information, since this is still being given away free at 800-4BVOCAL or 888-555-TELL. Instead, they've let users personalize: choose which stocks, teams, commutes, etc., they want to hear about, and in which order. And they've timed the browsing launch to coincide with availability of other monetizable, voice-activated services; chiefly voice-controlled messaging, dialing, and notification. We've asked, but preliminary take-up rates (typically \$4.95 per month) on each service are hard to come by.

But VAD's Not Bad

In pairing voice browsing's rollouts with voice dialing, the carriers and content companies have made the safest bet in speech. "The money-making speech applications still center in the carrier space, in three places: "Voice-activated dialing, directory assistance, and voice portal, with an emphasis on the first two," says Steve Ehrlich, vice president of marketing for Nuance.

Will Yapp, director of business development for network services at NMS Communications, sees voice-activated dialing becoming de rigeur among carriers, as well: "Speech vendors have the user interface down. The engines finally have some scalability to offer. The price still has to

come down significantly to scale in a network. OEMs are now starting to realize that this is achievable, so you're starting to get uptake from both them and the service providers on these services. It goes beyond sports, weather and stocks, into voice-activated dialing and voice-assisted messaging, which were not successful initially. MCI WorldCom is now rolling out a bunch of services. Things are now actually getting deployed in real network environments," he explains.

On the enterprise side, the easiest ROI for speech recognition is found in speech-enabling the old, prosaic call-center-offload IVR. Brokerage houses whose touchtone IVRs demanded callers to memorize ticker-symbols now recognize spoken stock and fund names. Ditto airlines, whose IVRs used to deal in (hard to memorize, easy to confuse) flight numbers, can ask callers for departing and destination cities and speak back confirmations. These new speech apps are robust and well accepted; and they keep expensive, high-turnover agents busy on money-making trades and reservations; not time-consuming information queries.

ROI For Voice Hosts (aka Voice ASPs)

Still on the enterprise side, the hosting proposition - you write it, you keep the database, we tap into it over IP and serve it over the voice network - is compelling and easily understood. It means pay-by-month, or by-transaction, instead of paying for (fast-obsolescent) IVR infrastructure up front. Better, it outsources IVR management and network monitoring, and (in theory) lets you scale. Scalability is enhanced and costs made lower by using ports to the fullest, sharing them among customers.

Great cases in point are described here under Telera and NetbyTel, but there are a growing number of success stories from other voice ASPs, as well. Voice hosts are trying hard to make their enterprise apps all but shrink-wrapped, promoting verticalized solutions that only need prompts and 800 numbers to brand them and databases (preferably XML-tagged) from which data is retrieved, converted, and spoken aloud.

At the same time, the number of companies identifying themselves as voice ASPs is increasing. The IVR and CT call-routing service bureau industries are proving fertile ground for born-again voice ASPs. "Some of them used to be doing enhanced 800-number routing, some pure touchtone outsourcing," says Stuart Patterson, CEO of SpeechWorks.

Repositioned IVR Platform

Likewise, voice ASP platforms themselves can - some of them, at least - be viewed as repositioned IVR servers with a VoiceXML interpreter added. That's why some "deployments" of "voice portal platforms" turn out to predate the emergence of VoiceXML itself; the app has pre-VoiceXML history on the NMS or Dialogic platform.

Some ASPs and their platform makers stress VoiceXML compliance, some don't. Either way, we're just beginning to see the benefits of this new markup language. In most actual deployments I've come across to date, the app is being developed largely by the host company itself. In other words, paying customers have not yet reached the web-hosting model, where webmasters large and small write their own applications,

and FTP them up to host servers. Hosts are still writing the apps and doing the integration work to end-user databases.

Clearly, there are benefits to outsourcing the whole job of application creation and deployment to the host, as opposed to developing VoiceXML and remote database integration expertise in-house. (See Chris Bajorek's "Dr. C" column for a take on issues pertaining to the voice-portal outsourcing decision.) But Nuance's Ehrlich says that the remote-developer model is slowly gaining. Nuance has 8,000 developers now on its developer network (6,000 of which, for all we know, are also registered with SpeechWorks' and the portal companies' online sandboxes as well). More significant, large data companies like Siebel and PeopleSoft are beginning to roll out voice-enabled versions of their applications, using their own programmers or consultants. "One thing's for sure about those guys. They know nothing about telephony or IVR," says Ehrlich. "They are completely reliant on someone who runs VoiceXML to host those apps and provide the telephony infrastructure."

Nuance has had a hand in advising Seibel on interface design, but Ehrlich reports that they've since taken it in house. PeopleSoft is mainly contracting out to a third-party Nuance developer called [JustTalk](#) (Ann Arbor, MI - 734-623.7954).

In addition to these big software names, "hundreds" of smaller companies are rolling out speech applications in various verticals. [Appriss](#) (Louisville, KY - 502-561-8463) is rolling out apps in the government space. They've long run a victim notification service, VINE, that tells crime victims and the general public the custody status of offenders; now VINE is speech-enabled.

"Many application companies, such as Appriss, are currently hosting themselves, but they don't want to be ASPs long-term," adds Ehrlich. Nuance is playing a matchmaker role here. Another specimen, X-Time, San Mateo, worked up an app in BeVocal's Caf (BeVocal's online development environment), using their VoiceXML interpreter, Speech Objects, and performance analysis. The app is an appointment scheduler, available via voice or web. One X-Time customer is a software trainer using the ASP to schedule classes.

In addition to selling, consulting, and educating the developer community in general, speech companies are encouraging the growth of voice ASPs by playing several matchmaking roles. In some cases, notes Ehrlich, this means finding developers a reliable platform host. In other cases, it means recommending streaming audio content houses such as iSyndicate, IT Network, ON24, Blue Wireless, and TelSurf Networks. These provide constant audio feeds of sports, news, traffic, and weather. Or they provide dynamic data for others to attach to their own recordings. Or they provide ready-made VoiceXML applications with which to access their content.

CONNECTING THE SPEECH DOTS

The research trail to end-user customers who are using speech on an ASP basis is long and connects at least four dots. Dot One is a core recognition company (a Nuance or SpeechWorks or Philips, typically);

Dot Two, a platform vendor (one of a myriad who deploy Dialogic or NMS or other telephony hardware, such as Lucent, General Magic, VoiceGenie, Verascape, and in VoIP realms, Cisco and others). Dot Three is the ASP: beneficiary of the "Dot One + Dot Two" value chain. Finally, you reach Dot Four, the tenant/ customer. In the writeups that follow, we'll examine each Dot, bypassing most of the usual product descriptions to home in on recent deployments and industry news.

Core Recognition Players

Speechworks

[SpeechWorks'](#) (Boston, MA - 617-428-4444) marketing department is certainly channeling the new zeitgeist, working hard to present hard-number benchmarks of veteran deployments (E*trade, United Airlines) and more recent customers.

In December, SpeechWorks added a TTS offering, Speechify, licensed and repackaged from AT&T, and a speaker-verification Dialog Module that uses either T-Netix or Veritel authentication algorithms. Later that month, they acquired Eloquent Technology, and with it, 12 languages of formant-synthesis text-to-speech. Formant-synthesis is a wholly computer-produced simulation of human speech. While perceived as less natural sounding, it is less taxing on computer memory. And one formant synthesizer can be put to work for multiple languages.

Manulife: Among the newer names, there's Manulife Financial, a retirement fund administrator. Manulife speech-enabled a common IVR - 401K query - but it made much more of the application self-serviceable than it had been before, simply because it could now recognize and act on 250 different 401K names. That's a clunky thing to do in a touchtone menu, with key-and-placement presses to indicate letters. If your caller doesn't know the fund code, it's not possible by touchtone at all.

For Manulife, SpeechWorks reports, the call abandonment rate fell by two thirds. Average call time fell from 12 to two minutes, and cost per call from \$4 to 40 cents. The recognition rate (of fund names) is over 99% on first attempt, on a system accepting 160,000 calls per month. While balance inquiries remain DTMF-driven for now, the application automated trades, portfolio rebalancing, and future allocations.

AT&T: At the same time that SpeechWorks' announced adoption of AT&T's core TTS to produce Speechify, AT&T, in turn, agreed to let SpeechWorks front its own internal call centers. SpeechWorks presentations now quote AT&T noting costs per call going from \$1.30 to 14 cents, and hold times from 90 seconds to zilch.

AOL-by-Phone: SpeechWorks also claims actual paying customers for a consumer voice portal. It reports that over 125,000 customers have already signed up for AOL-by-phone; initially offered for free to AOL subscribers, it is now a \$4.95 add-in, for service that can be personalized via HTML web.

Continental Airlines: In early April, Continental Airlines opened its flight-information ASR app up to the public, after years as an employee-only service. The English version was added to the existing InterVoice-

Brite touchtone IVR; the Spanish version was the first attempt at automation for Spanish-speaking callers. According to Continental's project manager Donna Schiffert, speech rec brought zero-out down from 30% to 15 or 16% of calls. Obviously, the ability to say the city of departure and arrival made the application much more friendly than one requiring knowledge of flight numbers. Equally obvious, it keeps the live agents busier handling reservations.

Only a week into initial launch, it's handling 12,000 to 14,000 calls per day during weekdays and up to 35,000 during exceptional weather conditions.

Nuance

Voyager was launched two years ago as [Nuance's](#) (Menlo Park, CA - 650-847-7839) VoiceXML interpreter and browsing platform, but Steve Ehrlich reminds us that this is targeted to carriers. For service-provider or enterprise voice applications, Nuance has the Voice Web Server, a platform of core recognizer, Nuance TTS, and VoiceXML interpreter. This is being licensed to Voxeo and other ASPs, as well as to Cisco and Siemens.

Nuance TTS is new: Called Vocalizer, it's a blend and a tweaking of Fonix's, Lucent's, and several other third-party speech synthesizers and will be released this quarter in nine languages. Vocalizer's recognition engine, with Version 7.0, comes in 23 languages.

Ehrlich can rattle off a list of recent deployments, particularly in Japan, Taiwan, and Brazil, with brokerages and carriers. One of the most interesting, from a San Francisco company called Telespree, retails a disposable, incoming-only cell phone with just one button on it. Relying entirely on voice dialing and network storage of personal phone books, it's due to be sold through wireless carriers, packaging pre-paid minutes together with replaceable battery.

Phonetic Systems

[Phonetic Systems](#) (Burlington, MA - 781-270-4123), known for its million-plus-name recognition engine, still does about 100% of its business with Phonetic Operator, the name-recognizing auto attendant. It's an obvious ROI, relieving receptionists of the basic phone-answering task. So obvious a payback, in fact, that most speech vendors have also come up with name-recognizing auto attendants. Phonetic Systems still holds the record, however, for speed of search on largest databases, with a patent on its phonetic search algorithms.

Robert Miele, director of product management, reports that customers have used Phonetic System's SDK to craft variations on Operator; one brokerage has linked the auto attendant, for password-holding callers, to an automatic out-dial to that employee's paging service. It makes particular sense in huge companies of 20,000 to 50,000 employees, using multiple paging services. Large Manhattan financial houses want to build in sales-force access to quarterly directories, saving on printing costs and exposure to misuse. Toward this end, Phonetic Systems is looking into adding RSA authorization and voice verification from [T-Netix](#) (Englewood, CO - 303-790-9111) or [Per Say](#) (Woodbury, NY - 516-

In an effort to encourage use of the search engine in common data retrieval apps, Phonetic Systems also has attached some impressive and plausible numbers to the call center-offload proposition. Their hypothetical ROI shows a savings of \$3,200 per shift in employee time for a 100-seat call center, where each agent averages four minutes in talk time per call, half a minute in wrap time, 12 calls an hour, eight hours per shift, and costs the employer \$20 per hour. Automating just 20% of routine calls - say 1,920 calls, or 160 employee hours - yields this startling per-shift number.

Phonetic Operator 4.0 is Phonetic Systems' latest release, with new support for LDAP directories.

Philips

[Philips Speech Processing](#) (Dallas, TX - 972-726-1200), makers of the Speech-Pearl recognition engine and the SpeechMania engine-plus-development environment, is delighted to talk about ROI, perhaps because they never beat the portal drum too loudly in the first place. Their money-making apps are not flashy, but, says Tim Walsh, vice president of sales and marketing, they're the voice-activated dialing, voice-directed voicemail and unified messaging services that we now see the portal pushers adding, simply because that's what market research and practice shows people pay for. They're also the recognizer behind a lot of CLEC directory assistance, much of it through their Preferred Voice partner. They have a new customer in Syracuse, NY, USA Direct, a VoIP carrier with speech servers colocated with Sonus switches. But it's not deployed yet.

Not to say that Philips isn't in the portal business at all; in August the company added KG Telecom in Taiwan, with the world's first Mandarin voice portal, to its other portal name, the service run by Omnitel in Italy. Ms.600 from KG offers 18 services, including financial and traffic updates, taxis, airline reservations, pizza orders, and movie reservations. Intecs Information, a systems integrator, developed the portal using Philips' SpeechMania development and recognition platform.

IBM Voice Systems

Last October, [IBM Voice Systems](#) (West Palm Beach, FL - 914-642-3000), told us about their Websphere Voice Server, the Via-Voice-recognizing, VoiceXML-interpreting add-on to the Direct Talk IVR platform or the Cisco Voice-enabled router.

Western Connecticut University: Asked for customers, IBM comes up with Western Connecticut State University, where students are entering work hours on a VoiceXML application residing in the WebSphere application server and served by the Direct Talk voice server. The recent announcement of Websphere compatibility with Dialogic's Voice Portal Platform was followed by a deployment announcement in China with portal Tom Voice. It doesn't use VoiceXML as a programming model, however. Bank of Scotia is offering a suite of banking applications that work on web, IVR, and are now being enhanced with speech-recognizing stock quotes and phone banking.

T. Rowe Price: Brokerage house T. Rowe Price, an IBM Via Voice user since 1999, is about to launch a natural-language understanding (NLU) application pilot to manage (or so they say, at press time) five of their customers' employee plans. The app will let customers on the retirement side of the business use speech to get account balances, quotes, and performance data.

Here, NLU means an application that, for example, understands what "that" refers to; as in, "what's the performance of that (previously asked-for) fund?"

Callers will speak social security numbers and PINs, get account balances automatically, and then be able to ask for more details in a free-form way, as in, "What's my balance in the international stock fund?"

"If they ask for a statement, we'll ask them for date-ranges," says Tom Kazmierczak, vice president of business operations development for T. Rowe Price. If callers are stymied by the lack of precise menus, they'll hear helpful prompts.

Brokerage houses, under increasing competitive pressure for corporate retirement plan accounts, are building in self-service apps to maintain promised levels of customer service. Speech rec becomes vital here, points out Kazmierczak: We have over 70 mutual funds now. If you don't know the code, it's a laborious process to have to touchtone through prompts."

To date, account changes still rely on touchtone entry, although that is due to change. About half of T. Rowe Price's account queries use IVR, 35% use website, and 15% use live agents. Currently, web-based interactions show much more transaction completions than IVR; T. Rowe Price is hoping to see speech technology bring IVR more into line with web.

"The trend of the last two to three years in employee benefits IVRs has been a declining retention rate and an increase in zero-outs, because the plans have become more complex," notes Kazmierczak. Speech technology, it is hoped, will reverse that trend.

T. Rowe Price's production platform is IBM's Direct Talk for AIX, checking databases on IBM RS 6000. Direct Talk uses a proprietary Artic960 card to communicate with a lineside T1 on the company's Nortel switch. Via Voice NLU servers are on the RS6000. The system as a whole supports 450 ports.

On a high level, the app system code falls into four parts: Binaries provided by IBM for Via Voice and Direct Talk; modifiable forms written in TCL (pronounced tickle); modifiable Direct Talk custom components, written in the Direct Talk standard state table language; and back-end components written in C, Java, and Cobol using DB2 as a database platform.

T. Rowe Price's pilot employer customers, while receptive to new technology, are geographically diverse. "We're taking 60 to 90 days to collect data and fine tune," says Kazmierczak, "and we'll allow them to make transactions (in addition to queries) before we roll out completely."

Platform Vendors

Converse Network Systems

[Converse Network Systems](#) (Wakefield, MA - 781-246-9000), with an installed base of over 360 service providers, serving more than 300 million subscribers in over 100 countries around the world, is happy to make the case for in-network portal platform. They've integrated their enhanced services platform with a voice browser and dubbed it Tel@Go. "The details of integration with switches, SS7, billing - these are not small or inexpensive issues, and we've done them already with our voicemail systems," says Michael Krasner, vice president and general manager of Converse's speech portal division. "The kinds of things we're doing in Intelligent Networking and release-trunk linking is a great savings in ports. You can't see those economies if you're platform is not in-net."

The only "portal" customer Converse can name to me at this point, though, is Sprint PCS. And here, the voice portal applications will be provisioned through HeyAnita. In the future, says Krasner, their browser will eliminate the extra port by retrieving VoiceXML applications over IP. And in the meantime, the company is racking up carrier customers for unified messaging and voice dialing in North America and abroad, where Philips is a stronger competitor to Nuance for core recognition.

The Sprint Voice Command application, an option on their PCS service, will soon be the largest deployment of speech rec in the world. Subscribers upload their Sprint PCS dialing directories through Palm Pilot or Outlook. The Sprint services are realizing an average \$10 per month apiece, of which voice-activated dialing is one option. They're also seeing higher retention rates, since loading one's Outlook directories to the service over the web is something of a customer entangler.

Intel/Dialogic

[Dialogic's](#) (Parsippany, NJ - 973-993-3000) voice portal platform goes from a 12-port developmental version to 96 ports in a 2U chassis. It needs only two JCT network interface cards to arrive at the 96 ports of speech rec, costing \$175 per port. Before the JCT's arrival, says Dialogic's Tim Moynihan, you needed five cards - three Antares and two network interface cards - to cover the same volume of traffic, and the solution retailed for \$300 per port. By about a month from press time, the same horsepower will take up one quad-span card in a 1U chassis, for \$100 per port.

Moynihan reports that SpeechWorks and Nuance have had to make some changes to their core engine to support Dialogic's CSP (Continuous Speech Processing) architecture, which puts activity detection, echo cancellation, and pre-speech buffering on the cards, freeing up the processor to perform faster recognition. Speech vendor APIs have been tweaked accordingly. Nuance and SpeechWorks have been joined by others announcing CSP support: IBM, Philips, England's Vocalis, Germany's Temic, Lernout & Hauspie.

Dialogic reports voice portal platform adoption in the Far East, with Mandarin and Cantonese applications in Hong King, and HeyAnita applications in Korea. Dialogic hardware, of course, powers several of

InternetSpeech

Making it to our attention right at deadline is a company called [InternetSpeech](#) (San Jose, Ca - 408-360-7730) that, like VocalPoint, is also aiming at voice-enabling HTML sites with no recoding of content. Their difference, says CEO Dr. Emdad Khan, is a true open browsing model that will voice-browse any website, not just one rebuilt for a "closed" voice portal. His netECHO server software, aimed at ASPs, ISPs, and telcos, also comes with email-reading and key-word Internet searching (piped through Google or AltaVista but customized for audio) and - most ambitious but not yet integrated - language-to-language translation. Pilots are in progress in the U.S., Japan, and China.

Dr. Khan demoed a voice-browsing session to a URL that we picked: nmss.com. We spelled the URL to the system, using a quick-reference card of "N as in Nancy" hints for letters that might be misrecognized. We got to the site and L&H RealSpeak did a good job of reading Natural Microsystems' home page. But no matter what the TTS, these demonstrations always remind you that the audible web is best for small nuggets of information. Knowing this, netECHO has a highlighting feature that should let the voice browser zoom in on just the desired parts of a site.

Lucent Speech Solutions

We reported that [Lucent Speech Solutions](#) (Naperville, IL - 630-979-7742) had worked the VoiceXML 1.0 spec into its high-density Speech Server, and was very actively promoting both the standard and the server to carriers and ISPs. See last December's story for details on the platform. We're waiting for announceable customers.

NMS/HearSay

NMS Communications (Framingham, MA - 508-620-9300), formerly Natural Microsystems, initially released its HearSay Voice portal platform without a VoiceXML interpreter, leaving that choice up to the developer. What made it different, then, from any other voice-enabled IVR server using AG 6000 (for PSTN) and/or Fusion (for IP media service) boards? API, says Will Yapp, director of business development for NMS. They'd seen some customers approaching speech-enabled IVR from the speech technology side, using speech company development tools; they'd seen others developing through NMS' own Natural Access, and a third group writing their own abstraction layer over both speech and telephony APIs.

HearSay was built to put the speech rec APIs under Natural Access, making it easier for those who were used to the NMS look and feel to incorporate speech.

"In the first version of HearSay," says Yapp, "We put Nuance's speech APIs as a service underneath Natural Access. Someone building an app could get access both to the Nuance APIs and to the whole suite of telephony APIs, with a seamless events model, saving them from writing their own abstraction layer."

Subsequent HearSay releases have integrated SpeechWorks core recognition, but in a way that leaves SpeechWorks' API more visible.

This is a strategic decision to attract customers who have used SpeechWorks only on Dialogic boards, and therefore have no previous familiarity, and no speed-to-market to gain, by folding development under Natural Access. Going forward, HearSay will submerge the SpeechWorks API back under its own APIs and also incorporate other rec vendors as well, starting with Philips, with an eye towards the European marketplace.

The latest news on the HearSay front is NMS' acquisition of Boston-neighbor Mobilee, first known as a voice-portal service. NMS will bolt Mobilee's own VoiceXML interpreter and audio streaming functionality onto HearSay, producing a "complete phone-to-web" platform to be called HearSay SoftServer.

The new offering packages Voice-XML source code and a reference platform with a suite of VoiceXML applications developed by Mobilee: voice portal, voice-activated dialer, email-by-phone, and instant messaging. Driving home the platform's plug-and-play proposition to service providers, NMS will also offer an optional service contract, to include a way to manage third-party audio content that can be integrated with a personalization engine.

The Mobilee acquisition puts NMS in the ASP business as well; the release notes that Mobilee will continue to support its voice-tenant customers, such as the Lycos speech portal.

Verascape

[Verascape](#) (Oakbrook Terrace, IL - 847-919-0873), a spinoff of veteran CT service bureau Vail Systems, has made a strong pitch to ISPs and telephony service bureaus with their scalable Verascape platform, launched at CT Expo in March. President Mil Ovan reports that they're now in trials with Oracle (mighty synergies with an XML-ified database suggest themselves, here), and with an IVR service bureau in the Bay area exploring their route to VoiceXML enablement.

The three-part Verascape platform comes integrated with ASR and TTS, interprets VoiceXML, and uses SIP VoIP protocol. Verascape [Call Director](#) performs location, configuration and management, as well as call routing and load sharing among multiple concurrent applications. The other two components, independently scalable, consist of Verascape Speech Server for ASR and TTS resources; and Media Gateway, for telephony interfaces to the PSTN. Clustered, the Verascape platform can scale up to 24 racks and over 65,000 PSTN ports.

Product launch for Verascape's platform was expected first quarter 2001. Ultimately, the prevalence of SIP-based IP networks is expected to lessen or remove the need for the media gateway piece.

Each Verascape Media Gateway handles up to a DS3 of voice traffic taking in PRI ISDN; line cards are Audiocodes'. Initial ASR was Nuance's but they pledge vendor agnosticism; furthermore, they've since announced integration with SpeechWorks' Speechify text-to-speech as well. Initial release will be based on Solaris X86 OS.

VocalPoint

VocalPoint is both voice platform and ASP; (see under ASPs, following).

VoiceGenie

[VoiceGenie](#) (Toronto - 416-736-0905), a spinoff of Array systems, is a software platform vendor with a strong voice in the VoiceXML chorus, touting 100% VoiceXML 1.0 compatibility in their middleware and agnosticism in their underlying implementation of ASR or TTS. They're also promoting developer courses and professional services to voice-enable (XML-ified) corporate databases.

They sell the VoiceXML gateway, which runs on a UnixWare dual-Pentium machine and includes the interpreter.

The company is generating a steady stream of press releases on new products and marketing and technology partnerships: In March, they announced SpeechGenie, a VoiceGenie VoiceXML interpreter with SpeechWorks ASR and Speechify TTS. In April, they announced VoiceGenie VoiceXML Gateway 5.0, a 100% VoiceXML-compliant gateway for both PSTN and VoIP traffic across both H.323 and SIP protocols, with Q3 availability.

In an effort to attract more developers, they also added a new lure to their developer website: VoiceGenie GenieTracer is a PC-based run-time version of the platform, for running and debugging apps without involving true speech ports. Available for download since May 1 for \$100, the GenieTracer consists of a GUI, Log Checker, Tracer, the interpreter itself, and related components. It lets developers enter the remote or local URL of a VoiceXML app, interact with it, step-through, jump, trace and monitor its execution. Instead of hearing TTS or recognizing speech, however, the simulator shows prompts on the screen and accepts "ASR" input via keyboard.

Customers? In the works. AINS, a traditional IVR system integrator in Washington, DC, is using VoiceGenie to automate the application for commercial building permits in Montgomery County, Maryland. Using Nuance ASR, it's due for deployment in April. It recognizes street names in a basic form-filling application, and schedules appointments for on-site inspections.

[Audium Corporation](#), (formerly Phone2Networks - New York, NY - 212-609-1320) is working on building a business developing sales-automation applications around VoiceGenie's gateways, as well.

General Magic

A platform vendor with a past in speech recognition and applications is [General Magic](#) (Sunnyvale, CA - 408-774-4200), makers of the Portico personal assistant service. They've come out with their own Voice-hosting platform, the MagicTalk Voice Gateway, which provides integrated VoiceXML, telephony, media server, and speech recognition services.

General Magic divides its platform offering into two parts: the core telephony and speech technology horsepower, with C++ interfaces, embodied in the media and telephony server; and the VXML Dialog Engine, which speaks Java. ORB middleware sits between the two pieces. The platform currently supports both Nuance and IBM Via Voice

recognition; SpeechWorks and Philips' SpeechPearl engines will be added soon. SpeechWorks' Speechify TTS will be added to the AcuVoice currently supported.

In addition to performing recognition and synthesized speech tasks, the Magic-Talk Voice Gateway submits requests to web/corporate databases, retrieves output, and renders it in voice. One server handles up to three T1s, or 23 concurrent voice sessions. It is VoiceXML compliant, although General Magic readily admits that the nascent standard is ambiguous in places and incomplete in others. "We support what is clearly stated," says Saeed Khan, director of product management, "and supply extensions to functions not clearly specified." British Telecom, a customer, has done compliance testing with one of their gateways. Khan reports: "Running their apps through our gateway helped them bring their own BT's VoiceXML code into compliance."

An included VXML Debugger, with line- or condition-specific breakpoint settings, helps develop VoiceXML scripts, and an extension API lets developers add their own tags. Version 1.0 runs on dual Pentium IIIs at 700 MHz and Windows NT SP6, with Dialogic D/480 JCT 2T1 or two 2 Dialogic D/300 cards. A deployment monitoring tool, demos, sample code, CDR generation, and documentation come with the package.

The platform sells per-developer licenses, at \$3,000 per seat, and per-deployment licenses, for \$25,000 to \$50,000 per T1, based on volume.

OnStar: Asked for deployments, General Magic comes up with the OnStar Virtual Advisor, a portal accessible by pressing the blue OnStar button installed in most 2001 model GM cars. Speech GM worked with automaker GM for over a year on the voice-enabled parts of their roadside assistance call center. General Motors has programmed the cell phone embedded under the button to ring up the service. The Virtual Advisor can give news, sports, weather, stock information, and tailors its reading to web-entered customer profiles. "The OnStar application had to be hugely scalable," said Kathy Layton, CEO of General Magic. "They're expecting millions of users," and GM (the auto) has licensed the app to other automakers, including Saab.

ASPs

NetbyTel

[NetByTel](#) (Boca Raton, FL - 561-237-0950) is a good example of voice ASP offering applications that need little more than database hookup and prompt recording. Their "Voice Commerce modules" are orderable by very specific function: Location Finder, Order-By-Number, Order-by-Name, Literature Request, Order Status Query, Delivery Status, Customer Survey, Price & Availability, and Lead Capture. They are not VoiceXML-compliant to date. Their modules require minimum customization to link to client databases, and that customization is NetByTel's task.

Extremely ROI conscious (a "Speech Savings Calculator" on their site takes in your company's call volumes, hold times, call lengths and employee overhead to calculate potential savings from CSR offload and 800-number per-minute charges), they claim that Office Depot realized an

87% savings when they put order-entry on their speech-enabled system. But those figures must be examined; that particular IVR was limited to existing catalog customers. Products have to be ordered by number, limiting the recognition task substantially over an app that might have to recognize multiple 100-word vocabularies of office products.

Mark Group: The Mark Group, a direct marketer of clothing and home decor items, started their order-query deployment with NetByTel on March 5. They produce three different catalogs - Boston Proper, Mark, Force and Strike, and Charles Keith - and associated websites for different goods and demographics. Each has its own 800 number for automated queries. This is not the same number, stresses Scott Bryant, Mark Group's vice president of operations, that customers use to call in orders from the catalog or websites. He does not want to surprise customers with automation, if they're used to speaking with live agents.

The Mark Group has given their automated query app a pleasant personality, branded her "Annie," and has only begun to publicize the service in package inserts and addendums to email marketing messages. "Annie" gets callers' order numbers (an alphanumeric recognition task) and also gets specific item numbers within the order; she reads back shipped date and carrier. She's very good at recognition, says Bryant. (If item numbers are limited to specific orders, this would shrink the grammar of possible items and lighten the recognition load considerably.) Annie also manages to convey a sense of humor and patience if you read her the wrong number.

Sounding like the true convert, Bryant says that "Speech recognition technology has progressed so much in terms of conversational flow; we're very impressed by how lifelike she sounds," he says. The same character handles queries for all three branded applications. If things continue to go well under full-ramp-up, he's considering adding a catalog order-entry app as well.

The deployment took two and a half months "from first sit-down," says Bryant. In addition to customizing the app for Mark Group, NetByTel worked with their MIS staff to attach XML tags to their DB2 database, running on an AS400.

Real-time data is sent through an Internet VPN using Geneva Message Queuing/Level 8. Mark Group incurred no additional expenses for this project, since it piped data to NetByTel's NOC on the same T1 they use for their HTML websites.

With just a month under their belts and no promotion, they're seeing 20% of non-order calls (a mere 35-50 per day) going to Annie; they're expecting that number to grow to 50% after online and direct mail promotion.

How did they find NetByTel? They were in the neighborhood, and advertised their service in a local newspaper.

NetByTel charges by the completed call or by the minute. The underlying speech recognizer here is SpeechWorks, running on 13 servers and three DS3 connections. Telephony resources, pooled among all clients, are

Intel/Dialogic's.

Telera

[Telera](#) (Campbell, CA - 408-626-6826), is another early entrant in the voice-host/ASP marketplace. They've gotten lots of media attention for their distributed platform and combined VoIP/LD resale/voice host offering. A Qwest IP network ties together Telera's voice-server POPs, which terminate 800-number traffic. Once at the POP, voice and application data flow over an IP network and thereby cut voice tenants' long-distance and on-hold charges. Telera can also forward calls to client's call centers.

Synapse Group: At least some Telera customers are actually keeping their own application servers resident. Case in point here is Synapse Group (Stamford, CT).

Blue Wireless: Another Telera tenant is [Blue Wireless](#) (Burlingame, CA - 650-552-9400), a developer of personalization software for wireless telcos. Blue Wireless hosts the [application logic](#), which learns end-user preferences for sports, news, and other types of content. It also manages streamed audio content, checking it for freshness, and taking care of the FTP or HTTP retrieval, caching, and indexing.

BeVocal

Last December, [BeVocal](#) (Sunnyvale, CA - 408-907-3200) announced agreements with Sprint PCS and Qwest Wireless to provide voice portal services to their subscribers. But the Sprint announcement turns out to have been premature. "Never announce a trial," says the PR person for HeyAnita, which ultimately won the Sprint contract. But as noted in our intro, the last word is far from spoken in the carrier voice portal story.

"Sprint's long-range strategy is neither BeVocal nor HeyAnita. They're still trying to figure out whether they should own or partner," comments Nuance's Ehrlich. Early to supply voice-activated dialing (using Nuance ASR and an IBM Direct Talk IVR server developed by InTouch, which was itself acquired by Comverse), Sprint's most likely partner, says Ehrlich, are the AOLs or MSNs or Yahoo!s who can get them applications, content and most important, subscribers.

The problem with carriers is that they don't have content partners and don't know how to write applications. AOL and MSN have both and have loyal subscribers, besides.

Even BeVocal's co-founder Amol Joshi concedes that long-range, networks will want to bring their voice-service platforms into their own networks. "In general, carriers and the North American market want hosted solutions from BeVocal first, to pick selected cities, prove it works, and then, depending on how much infrastructure they own, an how much capital expenditure they can tolerate, as they get to a certain call volume, they want it in house." They don't want to pay intra-LATA rates from BeVocal back to the network every time someone voice-dials."

Recognizing this, BeVocal has started to productize their hosting platform for sale to carriers to bring into their networks. Joshi says that trials are in

progress with carriers. Consumer portal services, on the other hand, can terminate with one dial and make more sense outsourced.

Joshi has two points to make to carriers: One, that BeVocal's TMS (Total Messaging Solution) will do all the fancy voice dialing, email reading, and notification applications while still routing to a preexisting voicemail server if one is already installed; and two, "The BeVocal platform is the same price point as a carrier-grade voicemail platform - about \$1,000 per port. For the same money, you can get a platform that can run different applications."

Qwest: The Qwest Wireless announcement, on the other hand, has taken hold. In March, BeVocal added messaging capabilities, voice dialing, voice authentication, and notification to the preexisting portal service on Qwest's wireless service. The carrier, after offering a free portal trial to its customers for several months, started charging \$4.95 for browsing. Starting in June, it will charge another \$4.95 for the voice messaging and voice dialing, storing up to 500 contact names and five numbers per contact. The carrier reports that they're "very pleased" with the uptake rate, which they will not specify further than "tens of thousands." The switch to charge is in line with industry trends overall (see AOL, Sprint, and Yahoo).

BeVocal's Amol Joshi is happy to pass on a Wireless Week report that among new Qwest subscribers, a full 50% are signing up for voice browsing, which accrue minutes-of-use in addition to add-on subscriber fees. Joshi also praises the job Qwest did in promoting these services in all advertising media. And it proffers the services through a quick-dial code: *999 (star www!).

How can some companies charge for sports, news, weather, traffic, etc., when Tellme is still giving it away for free? Well, by adding personalization, for one thing. This means specifying (through a website maintained by BeVocal but branded for Qwest) which stocks you want to hear about, what sports teams you follow, what traffic information you need, and the order in which you want this information. Down the road, says Joshi, it'll mean finding restaurants and automatically pasting their numbers in your address book. And the voice messaging isn't just voicemail: It's a system that, because it stores your email address, can email you driving directions if you'd rather read than hear them.

Under its own portal brand, BeVocal quietly continues to offer registrants free, non-personalized voice portal service. Joshi says that it works as a test bed: gathering user feedback, speech sampling and market data that can be put to work, once perfected, under Qwest's (or other clients') brands.

BeVocal has also come out with a late-April announcement that gives European VoiceXML developers access to the test telephony infrastructure of its BeVocal Caf data center and development environment. The VoIP link to Europe will be provided by the iBasis network. iBasis also plans to work with BeVocal to promote VoiceXML development in Europe, cosponsoring free training events, beginning in London in May.