LOOK WHO’S TALKING: VOICE-BASED SERVICES ARRIVE

By Kevin Werbach

Startups seeking to make Internet content and services available over the phone, with speech as the primary interface, have proliferated in recent months. Some two dozen companies have announced their existence, and one speech-recognition platform vendor estimates it has a hundred prospects planning to offer such services. There have been a string of major financing announcements by companies promising to transform access to information as radically as the Internet did.

Behind all this excitement is one basic concept: ubiquity. As voice-based services mature, they will help make the Internet a truly universal global communications medium.

There are more than a billion telephones worldwide, with more than 90-percent household penetration in most Western countries. Add to that several hundred million mobile phones, projected to grow to 1.4 billion by 2004. All these are network-connected by definition, and all are capable of taking speech as an input. Beside this, the current installed base of 200 million or so worldwide Internet users seems almost insignificant. Moreover, talking into a phone works in many situations, such as when people are driving or (unfortunately) eating in public places, where other interfaces and their supporting devices are impractical.

Given these factors, there may be no bigger opportunity in computing than connecting the vastness of the Internet with the most popular means of communication. And with the maturation of speech-recognition technology, this opportunity is becoming a reality. Companies such as United Airlines, Home Shopping Network and seven of the top ten US retail brokerage firms use speech recognition today to cut the costs of human call-center operators.

Beyond the enterprise environment, similar services are now appearing on the public Internet, in the form of a new class of companies known as voice portals. This is
happening at breakneck speed. “I think six months ago nobody had ever uttered the words voice and portal in the same sentence, and now they are doing it all the time,” says Ron Croen, CEO of speech-recognition software provider Nuance (see page 9).

Speak, memory

Speech was the first means of symbolic communication, evolving thousands of years before writing. It is the one communications form virtually every human can engage in,1 everywhere and all the time, because for local transmission it requires nothing more than sound waves. However, while the local transmission of speech is easy, interpretation of spoken messages requires a very powerful device such as the human brain. Our brains have evolved remarkably efficient language-processing circuitry over millennia. Linguists such as Noam Chomsky and Steven Pinker argue convincingly that we are born with the fundamentals of language pre-wired into our heads. Before we learn even a word of our mother tongue, allowing us rapidly to master the devilish nuances of syntax and semantics.

Recognizing words and understanding their contextual meaning are two different things (see Release 1.0, 1-99 for a similar discussion in the context of search engines). Given the sophistication of our own language apparatus, it should not be surprising that the state of the art lags so far behind the science-fiction vision of computers that respond perfectly to spoken commands.

The Net is a communications mechanism, but today we generally communicate with the Net through input devices designed around the needs of computers, such as keyboards. Keyboards can be slow, they are hard on the body over time and they take up space. A keyboard works fine for a desktop PC or even a laptop, but it becomes much less efficient when scaled down to a handheld device. Companies such as Research in Motion (RIM) have done a great job making usable mini-keyboards for short wireless e-mails, and the Stowaway folding keyboard for Palm devices is a wonder of engineering. However, even these smaller models are no help with mobile phones... not to mention when you're walking down the street or sitting in a car and simply cannot use any hand-based input device.

Another stage of convergence

There's another dynamic at work here: the integration of the telephone network into the Internet. Up to now, the Net has run on top of the telephone network to a large extent, but it has not competed with the services offered by the owners of those networks. In the traditional telephone world, voice is an end-product. The speech that passes over telephone wires and wireless connections is seen as separate from the network infrastructure that ensures it gets to where it's going.

With the emergence of the SS7 signaling protocol and intelligent networks (see Release 1.0, 12-99), the data-processing technology involved in managing the network began to connect to the services offered over the net-

---

1 Of course, some people have speech or hearing impairments that prevent them from communicating through audible speech. If for no other reason, speech-based interfaces will never totally replace other means of input.
work, allowing, for example, originating phone-number information to be passed to customer-service centers, and calls to be routed and billed in more sophisticated ways. But the integration of voice and data in the traditional telephone world stops at a certain point, because voice communications are still opaque analog signals.

In the Internet paradigm, bits are bits whether they encode voices, pictures or instructions between computers. The corollary is that those bits can be interpreted and treated as a set of network-enabled applications. Voice portals are significant because they move towards a world where ubiquitous, reliable telecommunications networks are as programmable and open as the Internet. These services make it possible to pick up any phone and interact with dynamic content and services, which over time will become increasingly personalized (see Release 1.0, 4-00 on the emergence of the “data soup” model). Today you can plug your analog modem into any RJ-11 wall jack and experience the whole Internet; tomorrow you’ll be able to do so without the modem.

Making Internet services available through ordinary telephones is also an important antidote to the “digital divide.” Even in the US, many people do not own or cannot afford PCs, and penetration levels are lower in most other countries. Hardware and Internet service costs, plus lack of familiarity with PCs, are barriers to many people’s adoption of the Net. Speech-based services promise to make some of the most useful features of the Internet available cheaply (or even free) to anyone with a telephone.

Of course, speech won’t supplant the more-established visual and hand-based computing environment, because there are many things speech isn’t good for. You can listen to something over a phone at the same time as you perform another function such as driving or reading a memo. However, you can’t interpret three voices at the same time, let alone simultaneously take in the dozens of options available through the links on most Web pages. (Imagine hearing the more than 160 hyperlinks on the Yahoo! home page read to you in order over the phone!) Phone-based services also have their own social and business-model elements (see page 23) that make them better suited for some uses than others. Consequently, speech will be one interface among several for Internet content and services.

THE VOICE-SERVICES LANDSCAPE

Speech-recognition technology has been kicking around in computer-science labs for some time, though in the past decade and a half it has become commercially viable on a broad scale. Some of this is due to increasing algorithmic sophistication, but the march of Moore’s Law is also a major factor. Speech recognition is an exceedingly complex real-time operation, and nothing helps more than having more processor cycles to throw against the problem.

Two classes of voice-centric consumer applications are now in the market: software designed for PCs and services intended to be used through telephones. The initial wave of excitement around voice recognition centered on PC-based dictation services from companies such as Lernout & Hauspie, IBM and Dragon Systems (which Lernout & Hauspie agreed to acquire earlier this year). These packages take voice input via a microphone and allow dictation directly into desktop applications such as word processors.

Release 1.0 23 May 2000
The technical challenge for dictation services is accuracy across an unbounded domain: you might say anything and the software needs to distinguish words in its dictionary from proper nouns and from non-printing commands such as "save" or "make this word italic." To improve accuracy, these software packages all encourage or require the user to train the recognition engine to his or her voice when first launched. They also incorporate natural-language processing algorithms to differentiate similar-sounding words such as "here" and "hear."

Phone-based systems have different challenges. The speech recognition must work under the less-than-ideal acoustic conditions of a telephone connection, which become even less ideal when mobile phones are involved. Because the services run off carrier-grade servers on the back end, they have more total processing horsepower to work with than desktop dictation software, but they must scale efficiently to support many users on the same server. Phone-based services need only support a limited domain of commands, rather than understand every possible conversation. However, they must support many users' accents and vocal quirks, because people don't want to go through an extended training period every time they pick up the phone to request information.

The other important distinguishing characteristics of the phone-based speech-recognition market are economic. As noted above, there are many more phones than PCs. Though wireless data services are taking off (see Release 1.0, 4-99), they are still only supported on a small percentage of phones. Moreover, even when mobile phones incorporate data services, such as e-mail and Web content, via the wireless access protocol (WAP) or some other mechanism, they still leave much to be desired. Only so much information can be supported through tiny four-line displays, and reading any screen is impractical in some situations such as driving.

A brief history of phone-based services

The first wave of phone-based speech services were personal assistants such as Wildfire (see Release 1.0, 10-94) and General Magic's Portico. These services feature an automated agent that offers voice-based dialing, voice-mail handling, unified messaging and related functions, generally targeted at busy professionals.

Wildfire launched in 1992 and has gradually gained subscribers, especially with recent wireless deals in Europe, but it has never come close to "growing like wildfire" as its name was designed to suggest. General Magic has also experienced some success, especially with its scaled-back MyTalk offering, but it has similarly not lived up to its promise. That hasn't stopped new competitors such as Webley from entering the market, promising that they will succeed where predecessors failed.

The earlier personal-assistant services were limited by the state of the art in speech recognition. Phone-based recognition engines until the past two years or so still required training to a particular user's voice

---

2 Recently, General Magic has made headway in the market for in-car services, taking a $15 million equity investment from General Motor's OnStar Division and signing a deal to be the interface for the OnStar Virtual Advisor service.
when going beyond the most limited commands. Service providers had to build much of their infrastructure from scratch, and it was originally not very scalable. Their business models generally included per-minute usage fees or large monthly charges, which deterred many users from signing up.

Also, though many business professionals find personal-assistant services useful, they aren't a killer app that drives massive adoption (at least not yet). Personal assistants require people to change the way they do things, which is an adoption barrier even if the new procedures are more efficient. By contrast, automated speech services that provide informa-

Standards are important to the growth of any new communications channel. The Web grew up around standard HTML, dynamic business-to-business services are coalescing around the extensible markup language (XML) and wireless data services build on WAP. In all these areas, proprietary approaches developed first, and in many cases are still in use, but standards catalyzed the rapid expansion of the market.

The primary standards effort for voice-based services is VoiceXML, which launched at the beginning of 1999 and released version 1.0 of its specification in March. VoiceXML brought together IBM's SpeechML, Motorola's VoxML and the Lucent and AT&T markup languages based on the PhoneWeb project at Bell Labs. The group has successfully expanded beyond those four companies to encompass virtually all the important players in speech-based services among its 140+ members.

VoiceXML is a specialized language built on top of XML. It allows voice applications to be built out of "documents" that define dialogues using standard markup tags. This approach will make it easier to create new applications, and will allow applications and users to migrate across different companies' platforms.

None of the commercial voice-based services we describe support VoiceXML today, through most companies say they plan to support the standard. Numerous companies are developing VoiceXML servers and browsers, though the familiar questions of when new features constitute valuable extensions and when they represent deviations from the standard have already begun to arise.

Even if VoiceXML is universally adopted, it won't prevent companies from developing unique value-added offerings, and more than HTML has precluded proprietary applications such as content management on top of the standard. The degree of adoption of VoiceXML will, however, affect the variety of voice-based services introduced, and the extent to which users can switch between services on the same call. Applications such as stock quotes, sports scores and access to e-mail are obvious starting points for voice services, and many providers are building such functionality. As the Web experience shows, however, even more powerful killer apps can arise from third parties or vertical-market specialists if the platform is sufficiently open.
tion or transactions may be seen as a completely new function, rather than forcing people to change the way they do existing things.

User expectations have also evolved thanks to the Internet. “The Web has been our best ally in developing the idea of remote self-service,” says Stuart Patterson, ceo of speech-recognition platform vendor SpeechWorks. Once users become accustomed to getting their sports scores or personal messages from a computer, it's not hard to think about doing so from a computer you call over the phone.

Newcomers everyone’s talking about

The plethora of emerging voice-based Internet services can be divided into three categories: voice portals, voice enablers and voice platforms.

In the first group, with the greatest potential reward but also the greatest risks, are the voice portals such as Tellme and Quack.com (see pages 11-16). There are two sub-categories: voice portals that rely on automated speech recognition, and those that initially use human operators, backed by automated systems in call centers. All these providers believe that the voice market will parallel the Web, in that aggregators and directories will capture the greatest share of traffic, though there will also be lower-volume, higher-value applications. Voice-portal proponents argue that the unique technical and interface elements will prevent existing Web portals from dominating the phone world, just as the portals superseded larger traditional media companies on the Web.

Voice enablers comprise the second category of phone-based startups. These companies deliver tools or hosted services so that Websites can offer speech-based services to their customers. The line between such companies and voice portals isn’t clear, because most of the portal companies also plan to speech-enable and/or host other sites as an element of their model. But given the demands of marketing and limited resources, the usual division between brand-building end-user services and behind-the-scenes tool providers seems likely to occur.

Finally there are the voice platforms, which support everything else. Here we find the speech-recognition engines, the basic application components and the network integration necessary to deliver voice-based services. Many of the big guys -- IBM, Motorola, Lucent, AT&T and Philips -- have stakes here, though the most aggressive suppliers to the new Internet-centric startups are two smaller pure-plays, SpeechWorks and Nuance (see pages 8-11).

To put it crudely, the first category is business-to-consumer (B2C), the second is business-to-business (B2B) and the third is infrastructure. In the Web world, popular interest and stock valuations cycled through these three categories. In speech, all three categories are hitting the scene at roughly the same time. We examine each of these areas below, starting with the platforms because they form the foundation for everything else.

All talk and no action?

The question is whether the same business-model dynamics that have caused many B2C sites to fall from favor will play out for voice-based services. One difference is that in the voice world there is a telephone carrier.
involved that often, especially for wireless, has a separate usage-based revenue stream that can be tapped. Moreover, the limited, targeted nature of voice-based interactions may make some models such as advertising and subscriptions more workable on the phone than on the Web.

Someone clearly finds these arguments persuasive. Tellme has raised over $50 million, and competitor BeVocal raked in $45 million in its second round. The VoiceXML working group (see page 5), which is developing standards for this industry, already has over 140 members, with more signing up almost daily. Of course, the market research folks aren’t far behind with their rosy estimates. In an oft-quoted study, the Kelsey Group predicts $5 billion in voice-portal revenues by 2005, plus $6 billion incremental revenue to hardware, software and network providers that serve those companies. All this when hardly any voice portals have even launched, let alone started generating meaningful cash flows.

Unsurprisingly, there is already a backlash to the wave of hype that has swept the land, raising the prospect that voice-based services will be passé before most consumers even hear about them. Sure, the demos are cool, argue the nay-sayers, but how will these services work under real-world conditions for millions of users? Moreover, even if the technology works, where will the revenues really come from? With all the competition, companies may have trouble maintaining subscription fees and persuading users to tolerate intrusive audio ads. Finally, even if one of the voice-portal startups may be the next Yahoo!, there certainly won’t be 20 voice Yahoos. A limited number of companies will survive the inevitable shakeout as independent entities, especially with established portals and major wireless carriers entering the market.

We believe there’s something real, and really important, in the new wave of voice-based services. But let’s not lose perspective. The early releases now launching are limited in functionality and their business models are uncertain. Once platforms are in place, though, expect to see rapid evolution and expansion in service offerings thanks to competition and the speech-enablement of existing Web and physical-world services.

VOICE INFRASTRUCTURE

At the highest level, the architecture of voice-based services resembles existing Web offerings. There are end-users with network-connected devices, tapping into content and services stored on remote servers. Upon closer inspection, though, there are important differences.

There is no browser on the end-user device, the phone, because it is ultimate thin client: able to operate without any software at all. Instead, if there is a “voice browser” it sits on the server side and acts as a client to the speech applications.

Voice content may be served from standard Web servers or databases, but in between the content and end-users are specialized software and hardware layers including speech-recognition server clusters, text-to-speech rendering and dynamic speech applications. Services that speech-enable existing Internet sites have an additional layer of intelligent agents that extract and reformat content. Figure 1 below provides a general overview of the major elements.
The platform players

Looking at the leading voice-portal contenders, virtually all of them use speech recognition platforms from one of two companies: SpeechWorks or Nuance. Lucent, Philips and IBM all have significant efforts in telephone-based speech recognition for enterprise and carrier customers, and given their resources and relationships they may still enjoy success in Internet-based voice services. At this point, however, the market is Nuance's and SpeechWorks’ to lose.

The two systems have similar roots in the research and academic worlds. Both can point to major enterprise customers that currently comprise the bulk of their revenues (including Charles Schwab, Fidelity, Home Shopping Network, American Airlines, Sears and UPS for Nuance; United Airlines, FedEx, E*Trade, MapQuest, BellSouth and Hewlett-Packard for SpeechWorks), and to customer wins in the voice-portal arena. Nuance completed its IPO in April; SpeechWorks is now in registration. Each sees a bright future, drawing analogies to the dynamic Website management platforms, such as Vignette and Broadvision (see Release 1.0, 9-98) that have rocketed to success in the past year.

Nuance and SpeechWorks each offer excellent recognizers as well as component technologies for building speech applications (Nuance’s SpeechObjects and SpeechWorks’ Dialogue Modules), along with VoiceXML support. In the most important areas -- accuracy and scalability -- both companies deliver reliable performance. When asked why they chose one over the others, customers rarely point to basic technology as the primary differentiator. Instead, they emphasize differences in the companies’ business approaches, strategies and product offerings.

SpeechWorks has gone further up the network stack, putting more emphasis into building and integrating its own speech applications and providing professional services so that customers can get up and running quickly. Nuance concentrates on platforms and tools, so that its customers have
more flexibility; for example, it provides open-source access to its foundation class of SpeechObjects.

Nuance

Nuance, based in Menlo Park, was spun out of SRI International in 1994. The company went public last month, the day before the stock market plunged dramatically.

Nuance has made voice portals an area of focus, and its software powers most of the best-known contenders such as Tellme and BeVocal. CEO Ron Croen sees Nuance as the market leader, and sees the market narrowing down as competitors such as Nortel drop out and companies such as Nuance continue to invest heavily in R&D. For example, Nuance's latest release of its core software, version 7.0, includes enhancements that Croen claims result in a 35 percent accuracy improvement when used in cars, which are an important future market (see page 22).

Croen sees three products cementing Nuance's lead in Internet voice services: SpeechObjects, VBuilder and Voyager. SpeechObjects are reusable software components, much like Enterprise Java Beans or Microsoft ActiveX controls but designed for the speech environment. Nuance has just released its foundation class of approximately 25 SpeechObjects, which any customer can use to build rich applications. VBuilder is what Croen calls "the equivalent of Frontpage for the voice Web." Scheduled to launch in the coming months, VBuilder simplifies the process of creating speech applications for the Nuance platform.

Voyager, announced in October and launching in beta shortly, is Nuance's voice-browser product. "Our browser is a killer app without being an app, because it's the voice enablement of all the content of the network," says Croen. Voyager includes a package of user interface conventions, such as bookmarks, back and forward transitions, personalization and hyperlinks; standard commands such as "help"; and a speech renderer that allows any VoiceXML content to be delivered through a voice portal. It also supports speaker verification, so that customers can be identified securely from voice prints and confirming information such as the phone numbers they called from (Home Shopping Network now uses this service so that customers need not re-enter information, including credit-card numbers, when making subsequent orders).

The goal of the voice browser is to give users a consistent experience regardless of the content and unique aspects of a particular site, much as the graphical browsers did for the Web beginning with Mosaic. "It's really voice dialtone, or intelligent dialtone," says Croen. "Each site you might call out to is discrete content, but the user enters the system through a browser in a consistent way."

SpeechWorks

SpeechWorks (originally known as Applied Language Technologies) was founded in 1994 with technology licensed from MIT's Laboratory of Computer Science. The Boston-based company has more than 200 employees, and filed for an IPO on April 20.
Tellme's announcement last year galvanized the voice-portal space, and since that time, says CEO Stuart Patterson, SpeechWorks has signed up ten to fifteen customers that see Tellme as their target. However, he views existing portals, now looking to speech-enable their offerings, as having “a bit of an unfair advantage” in the long run because of their established content, traffic, relationships and personalization data.

Patterson says that in the short term, most telephone carriers will choose between voice and WAP interfaces, if for no other reason than resource constraints. “If you’re a carrier really focused on WAP infrastructure, you’re too busy to make it speech-enabled at the same time,” he says. However, he predicts that before long companies will start looking at how to combine different interfaces to provide the best and most flexible user experience.

Patterson acknowledges that SpeechWorks and Nuance generally take similar approaches to the market. “Both of us are viewing ourselves like a Vignette or a Broadvision, where we deliver the technology, but other people host the services,” he notes. He argues, though, that SpeechWorks is ahead in key areas, having introduced its Dialog Modules two years before Nuance’s SpeechObjects, and bringing packaged applications and the...
open browser model to market first with its SpeechSite and Speech Portal products. SpeechWorks has also traditionally been more aggressive with professional services, he adds, in keeping with its emphasis on getting customers up and running quickly. Since SpeechWorks builds more of its customers' systems in-house than Nuance, Patterson asserts, it has developed better human-factors expertise and understanding of how to create effective interfaces. He adds, “Our goal is for our callers to hang up with a smile.”

Having systems deployed also will speed improvements in recognition accuracy, because the platform vendors can use real-world data (anonymously of course) to enhance their algorithms. “The more data we collect from people calling these systems, the better they get. Five years ago, with no public systems to speak of, it was difficult to improve the models,” Patterson says.

SpeechWorks is committed to VoiceXML, though Patterson agrees it’s early in the adoption curve of the standard. One consequence of standardization he sees is that VoiceXML renderers will become a commodity, though some companies may use voice browsers to lock customers into their custom extensions (shades of the Netscape-Microsoft browser wars). “Voice is a little bit less of a green field” than the Web was, he points out, but the degree to which de facto or formally constituted standards will gain traction remains unclear.

AUTOMATED VOICE PORTALS

“Voice portal” has become a common term to describe any speech-accessible Internet information service. Only some of these companies act like traditional portals, though, in aggregating content and services into a unified interface, with personalization and navigation features.

The first major voice portal in full commercial operation was BellSouth's SpeechWorks-powered Info by Voice offering, launched in the Southeastern US in January. However, Info by Voice is positioned more as an enhancement to BellSouth's directory services than as a standalone service, and BellSouth didn't design it to leverage the Internet in the same manner as the newer companies discussed below.

Tellme (...why I'm so great!)

Tellme is the most ambitious voice portal, and it also has the strongest Internet pedigree. Founded by Netscapees Mike McCue and Angus Davis, with a management team that includes well-known veterans from companies such as Microsoft and @Home, Tellme has quickly become known as a sort of Silicon Valley all-star team. Its Palo Alto headquarters in a converted printing plant is filled with homemade desks and loft beds (for those late nights...), an extensive collection of antique computers acquired on eBay, an old-fashioned British telephone booth and scads of Austin Powers memorabilia. (The Tellme service was code-named Mini-Me.) The company has raised $53 million from investors including Kleiner Perkins, Benchmark, the Barksdale Group and former Microsoft svp Brad Silverberg.

All well and good, but is there any there there? In the months since Tellme first trumpeted its existence last summer, competitors such as
Quack.com have appeared on the scene, claiming to leapfrog it. Tellme announced the controlled public beta of its service in April, and was immediately swamped with heavy usage. It is now scaling up its infrastructure, and plans full national rollout this summer.

A test of the beta service makes clear that Tellme has the deepest and most polished offering available today. It offers a smorgasbord of services including news, airlines, stock quotes, traffic, free two-minute domestic phone calls, weather, sports, restaurants, movies, horoscopes, blackjack, even soap opera reports. Some of these are self-contained applications, while others (such as airlines and restaurants) automatically connect outbound calls to the appropriate company. Tellme is also developing Web-based personalization tools, though in the current beta these are limited to selection of a list of favorite stocks.

The company has recorded a huge amount of voice content, such as weather reports, sports updates and the names of thousands of restaurants nationwide, complete with background music and even named announcers in some cases. At one point, the company had only 45 employees but 70 contract audio-recording professionals. As a result, most content in Tellme’s service avoids the computer-generated sound of text-to-speech engines. The whole thing feels like an entertainment experience, with each service area having its own character. Co-founder and chief tellme Angus Davis sees comprehensiveness as important to meet the needs of the broadest possible audience, which Tellme is targeting. “I grew up in a small town, and my family can use Tellme to find a restaurant just as easily in Bristol, Rhode Island, as in Manhattan,” he says.

In its early beta incarnation, Tellme’s Nuance-based speech recognition works extremely well on a wireline phone handset and quite adequately on a speaker or mobile phone, though as expected there is still some room for improvement. The $53 million question will be how well Tellme can scale its infrastructure once it goes to general availability nationwide, and how easily it can add additional content and services.

Tellme’s applications are built in VoiceXML and Javascript on top of the Nuance voice-recognition engine, with audio content served from standard Web servers. At launch Tellme will have two or three thousand ports of simultaneous capacity, though Davis expects by the end of the year to surpass the largest single installation of call-center capacity in the US (10,000 ports deployed by Citibank). Scalability is always a concern, Davis says. The company chose to build some functions in C code rather than use untested Nuance SpeechObjects, for example.

Tellme has been active in the VoiceXML standards effort, especially in the area of scripting support, though Davis acknowledges it’s early for standards to take hold. One benefit of the standards-based approach is that it makes it easier to create new applications using existing Internet content. Tellme is working on tools to speed this process. “We would like to make it possible for anyone to build something in VoiceXML on top of this service that we’re developing,” notes Davis.

The service is free to consumers and accessible through a (not yet public) toll-free number; revenues will come from a number of sources which Tellme is still evaluating. For example, service categories have sponsors who will pay to have themselves identified each time the category is
selected. Tellme is also experimenting with audio ads. And of course e-commerce transactions will generate revenue-sharing opportunities. Finally, Tellme plans to serve as a voice ASP, hosting services for other companies on its platform for a fee.

Davis says Tellme's business plan is not centered around splitting usage fees with wireless carriers, which creates incentives to keep people on the phone rather than deliver the most efficient service. "One of our core principles has been helping people get tasks accomplished as quickly as possible," he says.

**Quack.com: quick and convenient**

Quack's service launched nationwide on April 10 at 1-800-73QUACK. It offers weather, traffic, sports scores, stock quotes and movie information, along with personalization options. Quack's interface is more spare and businesslike than Tellme, with greater reliance on computer-generated text-to-speech output, though many of the basic services offer similar types of information.

Quack's founders argue that the speed with which they built and deployed a national service since the company's founding in August 1999 shows that their technology is more scalable and quick to deploy than Tellme's. Given this track record, they are confident they can improve and expand the service more rapidly than competitors. Quack also differs from TellMe and BeVocal (both Nuance customers) in that it uses SpeechWorks for its recognition platform.

Quack CEO Alex Quilici, formerly professor of electrical engineering at the University of Hawaii, sees voice-based services taking off because they are so easy to use, and because users can engage in another task, such as driving, while retrieving information. Moreover, because users can connect through any telephone, "impulse buys" may become more prevalent through voice portal interfaces. "The voice portals that can go very deep and wide in information, while still being simple to access, are going to win," says Quilici. He argues that services should allow users to specify their preferred information source, such as a particular critic for a movie review.

Personalization is an important element of Quack's service. Users can customize parameters such as the city and state for which they want movie listings, sports teams for which they wants scores and a stock portfolio. These preferences are set through the Quack.com Website, and then when using the phone service, the user starts the request with "MyQuack" followed by the service (stocks, movies, etc.) to request their information.

Quilici also believes that enabling other sites, such as portals, to deliver their services through a speech interface will be important, as will signing up carriers as a distribution channel. He argues that Quack's technology gives it an advantage when dealing with existing sites, because Quack can voice-enable Websites without them having to dedicate significant technical and personnel resources to the process. Quack has built a set of tools that its employees can use to quickly identify important pieces of information on a site (such as product listing and price information on an e-commerce site) and automatically generate an agent that extracts that content and turns it into speech.
The next two months will be critical to determining winners and losers, Quilici argues, as carriers, major Websites and infrastructure companies such as Lucent and Ericsson pair up with newer technology providers. Portals will be an important distribution channel, he says, though the most valuable partners will likely be wireless carriers: “One of the key costs in building a voice portal is communications costs. Bandwidth is free to the carriers. The most economical model is to provide it for the carrier and get a cut of the usage fees.” Currently, Quack’s service is supported by five to seven second audio ads. The company is an early adopter of SpeechWorks’ framework for targeted, interactive audio ads, called SpeechSpots.

BeVocal: location, location, location...

BeVocal, based in Santa Clara, CA, was founded in March 1999 by Amol Joshi, Mikael Berner, Kevin Stone and Steve Tran. Three of the four had worked together several years before at Panasonic Labs in Japan, developing consumer audio products and wireless systems. Seeing the growth of wireless data services in Asia led them to think about new ways of delivering Internet content and services, ultimately resulting in a focus on voice-based applications. The company, which now has 70 employees and expects to double that by year's end, raised $45 million in its second round from Mayfield Fund, US Venture Partners and Technology Crossover Ventures. It announced its service in January and plans a commercial launch in San Francisco in June, followed by a national rollout later in the summer.

Where most voice portals have tried to deliver as many services as possible, BeVocal has concentrated on developing a smaller number of more complex, unique offerings, along with an architecture that makes it easy for others to add new features and applications.

BeVocal's initial focus is on location-based and travel-oriented services. For example, it has built a driving-direction service that, when given starting location and destination, reads out exact directions using a combination of pre-recorded voice prompts, text-to-speech technology from Lernout & Hauspie and mapping information from MapQuest. BeVocal has filed for patents on technology to accurately identify place names and addresses in spoken input, and has also integrated its service with travel systems such as the Sabre airline reservation database.

Founder and vp of product marketing Amol Joshi explains the motivation for BeVocal's concentration on this area: "Location-relevant services are the most useful and frequently demanded by consumers, but there's no one company that can build all the applications. We want to use our strength in location-recognition technology to get our core offering out there, and then we want to enable a ton of other companies to build additional applications."

Location-oriented services are particularly valuable to brick-and-mortar companies since they want to drive traffic to their physical facilities. Joshi believes these companies will be the best initial customers for hosted voice services, not portals or e-tailers. “Most Web and e-commerce companies don’t have a phone business today, so the immediate value of a voice portal isn’t obvious to them,” he notes.
For example, BeVocal has developed a locator service that tells callers the closest Federal Express dropoff location and provides driving directions on request. Since physical-world services often involve transactions, BeVocal is less dependent on advertising revenues than other voice portals. For example, Fedex will pay a per-call fee for the dropoff locator service, because it drives customers to its package-delivery business. BeVocal also expects to generate revenues hosting speech-based services for companies that want their own toll-free number.

Joshi argues that “we are the open systems company in the voice portal space.” BeVocal’s service is built out of SpeechObjects (see page 9) and VoiceXML applications. Joshi says the company has the largest collection of SpeechObjects anywhere, more even than Nuance itself. And its cto Mikael Berner is active in the VoiceXML standards committee.

This month BeVocal announced a partnership with Nuance to create a SpeechObjects exchange, allowing third parties to license BeVocal’s objects and build their own applications. For example, a hotel chain could create a custom application for customers to locate the closest facility and book a reservation over the phone. BeVocal plans to charge a mix of up-front customization fees, recurring fees based on call volume and transaction-based charges depending on the specific situation.

BeVocal, like its competitors, has spent a great deal of time on consumer tests and focus groups to fine-tune its service. Joshi says the early focus groups reinforced the company’s decision to build a tightly-focused set of applications: “People that we talk to really don’t want to surf the Web over the phone. There’s a certain set of functions that they want access to.” BeVocal also learned that users interact with voice services differently depending on the context. For example, one person in a car generally wants to hear driving directions all the way through, but when there two people in the car, the one on the phone often wants to pause the service to relate the directions to the other person.

The company plans to deploy 12,000 to 16,000 telecommunications ports by the end of the year, and to add to that in 4,000-port increments. (For reference, the total national capacity of the popular Moviefone movie-listing service is about 4,000 ports.) Joshi believes that, unlike many Internet businesses, which must sustain losses for an extended period of time, BeVocal’s business model parallels traditional enhanced telecommunications services. In other words, once it recoups its initial infrastructure costs, it can make money on every call because its services tend to be transaction-oriented.

**Audiopoint: pinpointed services**

Audiopoint, which has only ten employees and has so far survived on angel funding, is among the least-known of the announced voice portals. With a beta service available to callers from Washington, DC since December, though, it was arguably the first to market. In early April Audiopoint launched its new version with expanded content and geographic coverage, including nationwide weather and traffic reports for twenty major cities. In addition, Audiopoint offers Web-based customization features. The service is available at (888) 38-AUDIO. Since it launched, Audiopoint has received calls from all 50 states, and it expects to reach one million calls by early July.
"It's very clear that speech is a very different technology, and it's a very different medium from the Internet and the Web in particular," says CEO Nick Unger. Just as traditional media companies failed to dominate the Web, allowing new entrants such as Yahoo! to build prominent brands, Unger sees a new generation of speech-oriented services competing successfully against existing Websites in the phone world.

But with this new opportunity comes new challenges, he notes: "The Web allows you this great opportunity to present a variety of information together. You can present eight concepts at one time and you can bounce between them. But on the phone you can't." Services therefore must anticipate what users want so as to cut down on explicit navigation options, without seeming to constrain user freedom.

Unger previously held several executive positions at interactive voice response (IVR) vendor PriceInteractive, and he believes Audiopoint's strength lies in its experience developing the technology and user experience of successful phone-based applications, married with its ability to integrate Web-based technologies such as personalization.

Unger says creating a good user experience over the phone is critical, but he sees this as a question of art rather than as a straightforward technical challenge. He worries that some of the more prominent voice portals will raise expectations about the accuracy of speech recognition and the possibilities of these services too high, so that users will become frustrated when their unrealistic expectations aren't met. "The trick with speech isn't just recognizing what somebody said; it's putting it in to context," he notes.

Audiopoint's service is free, with revenues from embedded audio advertisements. Unlike banner ads, which many Web users no longer even notice on a page, audio ads prevent you from doing anything else in the service while they are playing. This exclusivity also increases the likelihood that users will respond directly to offers to engage in transactions.

**HUMAN VOICE PORTALS**

Some voice portals don't use automatic speech recognition at all. Quixi and iNetNow combine human operators with Internet technologies, promising a more compelling service than would be possible with automated technologies today.

**Quixi: the value, not the technology**

In 1992, Quixi founders Evan Marwell and Robert Pines started INFONXX, which provided outsourced directory assistance services for wireless carriers. INFONXX now employs some 2000 people in five call centers, and it has handled over 300 million total calls. The company developed technology for "personal 411" services that would go beyond traditional directory-assistance lookups, but Marwell says wireless carriers were reluctant to deploy such a novel service until someone else proved the market and the technology.

"We figured out that the wireless carriers are an animal that doesn't want to be first at doing anything," Marwell explains. But the carriers
move quickly once convinced a new service is viable, he adds: “They want to be second; they never want to be third.” Believing in the opportunity, Marwell and Pines spun off Quixi as an independent company last October to launch the enhanced service directly to consumers.

Marwell think the automated voice portals have the right idea, but are pushing the technology envelope too far. “Voice is the right thing for today’s marketplace,” he explains. “Unfortunately voice recognition, especially in a wireless environment, doesn’t work well enough yet. There’s one interface that works today and it works really well...and that’s the person.” Consequently, Quixi has built a service that uses human operators to provide information to users, but it has deployed technology to make those operators more powerful and efficient.

Quixi users can upload their address books from applications such as Outlook or devices such as the Palm into the system. Then, when they call the Quixi number, they can ask the Quixi customer-service representative to put through a call to any person and have it connected without specifying the phone number. Quixi is currently a local call only in Los Angeles, but it plans to deploy other local dial-in numbers soon.

Quixi has developed proprietary call-center applications to manage this process, as well as unique call-routing and synchronization solutions. It is also building integration out to e-commerce sites, so that a customer can call Quixi and make a purchase over the phone with shipping and billing information automatically passed to the e-tailer’s systems.

The service targets the “time-famished” consumer: those users who have shown their willingness to pay for convenience by, for example, spending hundreds of dollars a month on heavy mobile-phone usage. (Marwell says 15 to 20 million Americans use mobile phones 800 or more minutes per month.) Initially Quixi charges $19.95 per month for its service, though Marwell says the company will offer several versions including some that involve only per-usage or transaction fees. “We’re finding that people are willing to pay to save time,” Marwell says.

Marwell believes that once Quixi proves the viability of its service in a real-world environment, wireless carriers will be quick to sign up to distribute it. “Our view is that the wireless carriers are all looking for a way to bring this whole mobile commerce thing to their entire base now, as opposed to having to wait until a lot of people have a WAP phone, or waiting until voice recognition works well enough to use that as a channel,” he explains. Carriers are particularly skeptical of voice-recognition systems, he claims, because early efforts to support voice dialing by companies such as Accessline and Intellivoice weren’t accurate or reliable enough. Moreover, even if voice recognition works well, it still may be inefficient for time-sensitive users, because navigation often is organized into hierarchical menus rather than the free-form conversation humans prefer.

This is not to say that Quixi will never incorporate automated voice recognition. The company will integrate other technologies that complement its service, such as WAP for confirmation of transactions and other information directly to a phone display. If there is customer demand for services such as information retrieval, Quixi will partner with voice portals that have deeper offerings in those areas.
Quixi, which is based in New York, currently has about 90 employees. That doesn't include the call-center representatives themselves, expected to be roughly 200-strong by Quixi's general launch this summer; they are outsourced to INFONXX and other companies using Quixi's technology. After running on private money from its founders, Quixi closed $27.5 million in late February from Flatiron Partners, Accel Partners, RRE Ventures, Kohlberg Kravis Roberts and the New York City Investment Fund.

iNetNow: your own personal searcher

iNetNow president Lenny Young sees his company competing directly with voice portals, rather than Quixi, which he categorizes as more of a personal information organizer. However, his company uses human agents to answer user queries over the phone for the same reason Quixi does: people provide better responses than machines.

Young worked at IBM and GTE several years ago before becoming an independent film producer. About a year ago, as he was becoming disillusioned with the film world, he came up with the idea for iNetNow. Young says when he was away from his office and needed a piece of information, he called an Internet-savvy friend who worked as a Website producer and asked him to track it down online. “What I started to realize was that if I'm talking to someone who knows the Internet, they tend to find things faster than I would find them, whether I'm at home or on the road,” Young explains. “The easiest way to get information on the road would be simply to talk to somebody.”

Based on this notion, iNetNow was founded in July 1999, launched this March, and currently has about 100 employees, of which 50 to 65 handle incoming phone calls or research information to put into the customer-service knowledgebase. It has been funded so far by private investors, primarily from the wireless industry, though now that the service is live Young is in discussions with VCs.

iNetNow customers, who initially pay $19.95 per month, can request any piece of information available online, ranging from weather updates and sports scores to more esoteric searches. Young says 85 percent of calls require some contextual information beyond what an automated speech portal could deliver. For example, a user might request a stock quote and then ask something like “what’s going on with that stock?” which the agent could answer by pulling up news headlines on the company or by doing further research. “A huge part of what people do online is search on search engines, and you can't do that on a WAP phone,” says Young.

iNetNow logs all its searches and is developing a human-edited knowledgebase to make searches quicker and more efficient over time. At some point, the company may license the knowledgebase as a standalone search engine, similar in some respects to the popular Ask Jeeves service. For now, though iNetNow is targeting the “information junkies” who have shown their willingness to pay for services that save them time and deliver value. The company is approaching wireless carriers and other partners, and anticipates delivering its service on a per-call basis in addition to the monthly subscription option.

Unlike Quixi, which has outsourced its agents to call centers, iNetNow currently uses only in-house personnel, which it can more easily train.
and manage. The company is building out technology to further speed the searching process for the agents, and also to store user log-in and password information so the service can return the same personalized information the user would get directly from the Web.

Indicast: voice-directed audiocasting

The voice-based services market promises to be as almost as diverse as the Web, with companies creating novel hybrids of existing Internet, communications and media services. For example, Indicast, a 25-person startup based in Carlsbad, CA, plans to launch a service this summer that combines voice navigation with audio content aggregation. “The concept from the beginning was to be able to pre-select what you were interested in, much like a My Yahoo! page, then to be able to deliver that in an environment that is real-world,” explains vp of marketing Kevin Nelson.

Indicast hopes to OEM its service to wireless carriers and others who already handle large numbers of phone calls. President and CEO Bob Osias, who spent 18 years in the wireless industry and joined Indicast last month with the closing of its first venture financing round, says the service will offer unique pre-personalization of audio content from sources such as Dow Jones, AP and ZDTV into something analogous to personal radio stations that people can listen to when on the go or in their cars.

Unlike AudioBasket (see Release 1.0, 4-00), which is designed for playback of specific “baskets” of content, Indicast users will be able to navigate and select additional information through a speech-recognition interface. Osias is reluctant to give too many more details at this stage, other than to predict boldly, “What cellular was to the landline phone, Indicast will be to Internet access.”

VOICE ENABLERS

The lines between front-end voice-service providers and back-end enablers are difficult to draw, because all the companies offering services directly to consumers also proclaim their intent to speech-enable other Websites. However, some companies are focused primarily on delivering tools and services to sites or to wireless carriers. These companies generally bet that most users will want existing content or e-commerce offerings speech-enabled and re-aggregated for phone delivery.

NetByTel: speech ‘R’ us

NetByTel president Paul Robinson says speech is a natural opportunity because, “It’s the most widely available channel.” He continues: “Everyone is talking about WAP and PDAs today, but the phone is already there. Most people are around the phone all their life.” NetByTel concentrates on speech-enabling transactional Websites, preferring to focus on e-commerce rather than content because it believes the economic models make more sense. The company has developed technology that it claims can quickly make any site phone-accessible, by pulling information directly
from the site in real-time. Initial customers include Priceline, Office
Depot and BigStar.

Ceo Neil Bernstein emphasizes that the user experience over the phone
will be different from what people are used to on the Web: “You’re not
ever going to use the telephone to surf the web the way you do today with
your mouse and keyboard. For the most part, when you interact with the
Web by telephone, you’re going to do so knowing what you want to accom-
plish.” Users want to retrieve a piece of information, such as a stock
quote, or execute a particular transaction, through a directed dialogue
rather than the more open-ended wandering prevalent on the Web.

NetByTel can speech-enable sites in three ways. At the simplest level,
its software agents retrieve information from the site over the public
Internet, just like any customer, and automatically deliver the necessary
information to users over the phone. The second option is for NetByTel
to have access to a private Website, so that it is not dependent on per-
formance of the e-commerce site over the public Internet during busy
times such as holidays. Finally, the site can provide NetByTel with an
XML interface into its underlying databases, allowing the smoothest and
most efficient integration.

In effect, NetByTel’s agents synchronize the phone-based service with the
existing Website. Then, having mapped the data, NetByTel delivers it to
users through a set of applications such as shopping carts and dynamic
question-and-answer interfaces. “Anything you can do sitting in front of
a terminal, we can do with our automated agents,” says cto Dewey
Anderson. It currently takes two to four weeks for NetByTel to speech
enable a site, though it hopes to reduce that period going forward.

Anderson built BellSouth’s speech portal working with SpeechWorks, which
also powers the NetByTel offering. BellSouth had been interested in
phone-based speech recognition for some time, seeing the opportunity to
dramatically cut costs of human operators, and it funded 20 percent of
the research budget for Professor Victor Zue at MIT, whose work eventual-
ly formed the basis for SpeechWorks.

NetByTel hosts the speech-based services for its customers, freeing them
of the need to deploy and maintain new forms of infrastructure. The com-
pany currently has 1,200 ports deployed, meaning it can support 1,200
simultaneous calls, and it plans to grow to 7,000 ports in the next six
months. “We’re the Inktomi of our space,” says Robinson. “We’re
enabling technology.” Though NetByTel shares facilities across multiple
customers, each customer’s data and applications are kept separate, and
each customer gets a unique dial-in number to market to its end-users.

The service supports both inbound and outbound dialing, so that, for
example, the system can automatically call customers to tell them that a
shipment has gone out or that they have been out-bid in an auction.
NetByTel charges its customers a percentage fee on transactions (typical-
ly less than ten percent); if the transaction doesn’t generate revenue,
such as customer-service requests, NetByTel charges on a per-minute or
per-catalog basis. To speed adoption, there are no up-front fees.

The company, based in Boca Raton, FL and founded in June 1999, is funded
by Chelsea Capital Partners, Mesco and Deutsche Telecom’s T-Ventures.
Talk2: lots of talk

Talk2 has taken an interesting marketing approach so far. It has run two full-page ads in the *Wall Street Journal*, in June and November of last year, making deliberately hyperbolic claims such as “[We] anticipate blasting Yahoo!, Lycos and Infoseek completely off the planet by 2001.” However, Talk2 has not yet launched any consumer services and its Website provides virtually no information.

It turns out the company has been in business since October 1998, started beta testing its service in December 1999 and anticipates launching its first live trial toward the end of the second quarter with “a major wireless provider in the Los Angeles area.” Talk2 also plans a standalone service some time this year. The company is based in Salt Lake City, UT, and currently has about 70 employees.

Talk2’s three founders come from the wireless and long-distance carrier world, with experience at companies such as MCI. Cto Darren Wesemann built the core infrastructure over nearly five years at Xanthon, a 20-person company that Talk2 acquired. The underlying platform supports the queueing, security and messaging functions necessary to deploy speech-based services, and Talk2 is building the application functionality necessary to deliver a complete service.

While Talk2 offers pre-packaged services as the voice portals do, the company focuses on making existing Internet content accessible over the phone. Most users have already chosen their favorite Websites and have personalized them to provide just the information they want; Talk2 president Dave Morton argues phone-based access should leverage rather than replace those selections. Says ceo Brian Charlesworth: “Everybody else is aggregating information and being a content provider, whereas we’re allowing you access to whatever content interests you. The sites that I used before are the same sites that I access when I’m on the road.” Moreover, working with established content aggregators frees Talk2 from having to maintain and update the underlying information databases.

Talk2 says it can voice-enable not just public Web content, but e-mail, e-commerce and other “critical communications” functions. The service supports voice links and voice bookmarks, so users can assign a name to a Website and go directly there through the speech interface, rather than having to speak an unwieldy URL. Talk2’s technology distinguishes text and links on a Web page, so that users can follow links just as they do through a traditional screen-based interface.

Charlesworth and Morton emphasize the scalability of Talk2’s system. They claim it has scaled beyond 30 million users in tests at the HP and Sun benchmarking labs. The Talk2 platform has been designed to be modular, so that components such as the speech-recognition and text-to-speech engines can be swapped out quickly.

The company has a strategic relationship with HP, which is also an investor. Under the arrangement, HP is providing hardware and also collaborating on technology development, such as integrating Talk2 with its e-speak integration technology (see *Release 1.0*, 1-00). The company is now raising a venture round, and plans to make more-specific announcements about its strategy in the next 30 to 60 days.

*Release 1.0* 23 May 2000
The broader market

Companies focused entirely on voice-based Internet services will ultimately be only one part of the market. The leading speech platform providers, Nuance and SpeechWorks (see pages 8-11), both see three other major customer categories: telephone carriers, car companies and traditional portals.

Telephone companies, especially wireless operators, need to find new value-added services as their pricing models for basic dialtone are threatened by competition and new technologies. “Every telco has the same business opportunity that the startup voice portals have,” says Nuance ceo Ron Croen (see page 9). “Our major telco customers are repositioning their whole businesses as voice portals.” Similarly, SpeechWorks ceo Stuart Patterson says that “the people who have brought us touchtone for the past 20 years know that if they don’t become speech platforms, they are going to be out of business.” Both Nuance and SpeechWorks are in discussions with carriers; expect to see pilot services beginning to launch later this year.

In-car services will provide automated information such as directions through built-in mobile-phone connections and global positioning system (GPS) receivers. GM’s OnStar service provides an early taste of this market; though it still uses human customer service representatives today, it is working with General Magic on an automated system (see page 4). The car companies will control access to this market, much as telephone companies own the networks through which users connect to the Internet, but so far they have recognized that others will be better at developing the content and services. The car environment, with its high ambient noise, poses particular challenges for speech recognition, but the platform vendors are focusing particular attention on this problem.

In the final category are the existing Web portals and e-commerce sites. All the leaders, including AOL and Yahoo!, have “-anywhere” initiatives to put their content on non-PC clients such as PDAs, televisions and wireless devices. And they have content, brands and aggregation skills that are equally relevant on other platforms. So far these companies have chosen either to build systems themselves or to work with enablers such as NetByTel, but there’s no question many will partner with or acquire standalone voice portals.

The rest of the pack

Several other startups have announced plans to offer services that allow users to, in effect, surf popular Websites over the phone. One could call these companies audio ISPs, though most also plan to provide back-end services to companies that wish to optimize their Web content for voice delivery.

Two that have announced plans are TelSurf and InternetSpeech. Both are SpeechWorks customers. TelSurf offers a free service at (888) TEL-SURF, currently only available to callers in the Los Angeles area but expanding nationwide, that delivers a wide range of Internet services, including
news, e-mail, instant messages and personalized MyYahoo! content. The catch is that users must fill out a questionnaire beforehand so that the audio ads that support the service are targeted to their interests and location. InternetSpeech, which will launch in the San Francisco Bay Area at the end of this month and nationwide later this year, takes a different approach, charging $29.95 per month for its consumer service. Ceo Emdad Khan says the company's advantage lies in its intelligent-agent technology that is able to parse Web and filter content to deliver only the relevant keywords through an audio interface.

TALKING THE TALK

Interface is everything

The biggest challenge for voice portals may be developing effective interfaces. Most people have experienced "touchtone hell," navigating through seemingly endless menu trees to find the right information (or not...) through keyboard-based interactive voice response (IVR) systems. Users looking for quick information retrieval and Internet-like convenience won't tolerate wading through such rigid interfaces, and they won't wait for all the options to be read out to them.

In other words, voice-based services will need to anticipate what the user is looking for and the possible responses at any given time. They will have to offer both an easy-to-remember set of universal commands and also a wide variety of optional ways to say things in recognition of the different ways people talk. Speech recognition platforms such as those from Nuance and SpeechWorks have evolved numerous features to make them more intuitive, such as "barge-in" (the ability to say a command without waiting for the prompt or content being spoken to finish), confidence scoring (distinguishing definite matches with user input from those in which the system has a good guess but wants to confirm) and ways of phrasing prompts that make users feel more comfortable.

Other aspects of user experience are less obvious. For example, Tellme co-founder Angus Davis says his company originally designed its service to provide much more information about menu options to new users than to more experienced users. The theory was that those less familiar with the service would need explicit prompting, while returning users would remember the commands. Focus-group testing suggested exactly the opposite approach. New users felt overwhelmed by all the choices, and wanted to hear only the simplest options when starting out. Experienced users were more interested in exploring and hearing all the choices.

Some interface elements will become more familiar and standard as people simply get used to them. Much of the way we handle speech is based on social conventions. When Alexander Graham Bell invented the telephone, it was not clear what someone should say when picking up the receiver. Graham Bell himself is said to have preferred "ahoy," which (we're glad) lost out in America to the now-universal "hello." The growth of mobile phones and other wireless devices such as pagers and personal digital assistants (PDAs) has created a new culture of behavior. Today, it is not at all startling to see people walk down the street with phones pressed to their ears, yapping away (though combination microphone/ear-
phone connectors still create the strange sense of seeing someone talking into space).

The only way for service providers to create effective voice user interfaces will be to test and refine their offerings repeatedly. This is the area where companies are likely to distinguish themselves over time, even if their initial offerings are similar.

OmniSky: wireless data services

At the same time as speech-based services are proliferating, other companies are introducing new forms of wireless data that operate more like the conventional Internet than the limited options available so far. These services will coexist with, and at some point likely merge with them.

The most interesting new wireless data service is OmniSky. Company president Barak Berkowitz, formerly general manager of Infoseek Disney's Go Network, unabashedly says OmniSky has taken a page from America Online's playbook. OmniSky sees itself as the enabling service provider for a true wireless Web. Since today few people have wireless modems, the company is bundling a version of the Novatel Wireless' Minstrel modem for the Palm V as part of its initial offering. OmniSky began with the Palm V because it's the most popular handheld model today, but it plans eventual support for other Palms, Windows CE devices, pagers and WAP phones.

Most current wireless data services operate on the notion that limited screen real estate and input devices necessitate controlled, proprietary services that package bits of Internet content into special formats, rather than the wide-open expanse of the Web. OmniSky allows users to go to any Website, though sites it has partnered with will have content more precisely optimized for display on a small black-and-white screen. The service also provides wireless e-mail and directory functions. It currently costs $299 for the modem, plus $39.95 per month for unlimited access.

OmniSky is based in Palo Alto, CA. It raised initial funding from 3Com Ventures and Aether Systems, and closed $75 million from News Corp. and PSINet in early May.

Speech links?

How, if at all, sites and services will connect with one another in the voice world is an open question. The Web was designed from the very beginning with standard hyperlinks and addressing mechanisms, so users could quickly jump from one page to another. Browsers have forward and back buttons built into their interfaces, and with cookies and frames, it's possible to maintain some degree of context or state when moving between sites. These features are responsible for much of the feeling of openness and freedom the Web provides, because any other site is just a click (or a typed-in URL) away.
Voice-based services have no such capability today. Each platform is an isolated island, even though users may be able to retrieve the same Web content through different voice portals. VoiceXML and the standardization efforts Nuance and SpeechWorks are undertaking suggest that bridges between services may develop, but by the same token competition between services and platform vendors may push in the other direction.

And just how should a “speech link” operate? Should it be a blind pass-through to another service, or should it be a more “framed” transition that allows you to preserve your settings and to back up to the original service? From a technical standpoint, how can calls be passed off across different platforms, with billing and telephony infrastructure costs apportioned in an appropriate way?

Today the battle among voice portals and similar companies is all about speed to market, establishing brands and signing up marquee customers and partners. It’s like the nascent commercial Internet in the first half of the 1990s all over again, only this time everyone knows there are millions of users and billions of dollars at stake. Soon the next set of questions -- about architectures, user experience, business models and platforms -- will come to the fore. One thing is certain: there’s a lot left to talk about.

---

3 This is an area both Nuance and SpeechWorks are exploring. SpeechWorks offers some ability to maintain context in audio advertisements with its SpeechSpots offering.
RESOURCES & PHONE NUMBERS

Nick Unger, Audiopoint, (703) 279-5180; nunger@audiopoint.net
Amol Joshi, BeVocal, (408) 748-8700; fax, (408) 748-8888; amo1@bevocal.com

Bob Osias, Ken Nelson, Indicast, (760) 438-5700; fax, (760) 438-5701;
rdosias@indicast.com, knelson@indicast.com

Lenny Young, iNetNow, (818) 734-9099; lenny.young@inetnow.com

Emdad Khan, InternetSpeech, (408) 360-7730; fax, (408) 360-7726;
emdad@internetspeech.com

Neal Bernstein, Paul Robinson, Dewey Anderson, NetByTel, (561) 988-5050;
fax, (561) 988-5092; nbernstein@netbytel.com, probinson@netbytel.com,
danderson@netbytel.com

Ron Croen, Nuance, (650) 847-7700; fax, (650) 847-7931; croen@nuance.com
Barak Berkowitz, OmniSky, (650) 473-9700; fax, (650) 323-6785;
barak@omnisky.com

Alex Quilici, Quack.com, (408) 747-7330; fax, (408) 747-7311;
alex@quack.com

Evan Marwell, Quixi, (212) 989-5310; fax, (212) 647-8545;
mobile, (908) 507-7519; evan@quixi.com

Stuart Patterson, SpeechWorks; (617) 428-4444; fax, (617) 428-1122;
stuartp@speechworks.com

Brian Charlesworth, Dave Morton, Talk2, (801) 924-8100; fax, (801) 924-
8101; bcharlesworth@talk2.com, dmorton@talk2.com

Mike McCue, Tellme, (650) 930-9099; fax, (650) 930-9101; mikem@tellme.com
Angus Davis, Tellme, (650) 930-9001; fax, (650) 930-9101;
angus@tellme.com

Release 1.0 is published monthly except for a combined July/August issue
by EDventure Holdings Inc., 104 Fifth Avenue, New York, NY 10011-6901;
(212) 924-8800; fax, (212) 924-0240; www.edventure.com. It covers soft-
ware, the Internet, e-commerce, convergence, Web applications, network
infrastructure, groupware, streaming media, enterprise applications,
wireless communications, intellectual property and other unpredictable
topics. Editor-in-chief: Esther Dyson (edyson@edventure.com); publisher:
Daphne Kis (daphne@edventure.com); editor: Kevin Werbach (kevin@edven-
ture.com); systems director: Scott Giering (scott@edventure.com); market-
ing manager: Joanna Douglas (joanna@edventure.com); office manager: Bill
Demeritt (bill@edventure.com); circulation manager: Natasha Felshman
(natasha@edventure.com); executive assistant: Irene Lawrence
(irene@edventure.com); assistant: Kara Holmstrom (kara@edventure.com).
Copyright 2000, Edventure Holdings Inc. All rights reserved. No materi-
al in this publication may be reproduced without written permission; how-
ever, we gladly arrange for reprints or bulk purchases. Subscriptions
cost $795 per year, $850 overseas.
**Release 1.0 Calendar**

**2000**

May 31 - June 2  *Internet & Society 2000* - Cambridge, MA. How opportunities and ethical dilemmas in the Internet age are changing our lives. Call (617) 204-4234; email is2k@harvard.edu; www.is2k.harvard.edu.


June 4-7  **MindShare** - Napa, CA. Second annual Jupiter Executive Forum. For more info, call Tara Donnelly at (800) 224-6054; fax, (212) 780-5382; tara.donnelly@jup.com; www.jup.com/events/mindshare.


June 12-15  **Telecoms @ the Internet VI** - Geneva, Switzerland. European ISPs and telephone companies get together. To register, call +44 207 915 5055; fax, +44 207 915 5056; www.iir-conferences.com.

June 14-16  **Global Forum 2000** - Paris, France. Discuss e-Europe opportunities with heads of multinational companies. For more info, call (212) 522-2525; fax, (212) 467-0498; fortuneconf@pathfinder.com; www.fortune.com/future/conferences/global_descript.html.


June 23-25  **Telluride Tech Festival** - Telluride, CO. Join the fun! Contact Scott Brown; (970) 728-7000; fax, (970) 728-7001; scottbrown@rmi.net; www.telluridetechfestival.com.


* Events Esther plans to attend.  # Events Kevin plans to attend.

Lack of a symbol is no indication of lack of merit. The full, current calendar is available on our Website, www.edventure.com. Please contact Joanna Douglas (joanna@edventure.com) to let us know about other events we should include.

Release 1.0  23 May 2000
SUBSCRIPTION FORM

□ Please enter my subscription to Release 1.0 at the rate of $795 per year in the U.S. and Canada. Overseas subscriptions are $850, airmail postage included. Payment must be enclosed. Satisfaction guaranteed or your money back.

□ Please update my address as listed below.

Name ____________________________
Title ____________________________
Company _________________________
Address __________________________
City __________________ State _______ Zip _______
Country __________________________

Telephone _________________________ Fax __________
E-mail _____________________________ URL __________

□ Check enclosed
□ Charge my

□ American Express  □ Master Card  □ Visa

Card Number ______________________ Expiration Date __________
Name and Billing Address ________________________________
Signature ____________________________________________

□ Please send me information on your multiple copy rate.

Please fill in the information above and send to:

EDventure Holdings Inc.
104 Fifth Avenue, 20th Floor
New York, NY 10011

If you have any questions, please contact us at (212) 924-8800;
fax (212) 924-0240; e-mail us@edventure.com; www.edventure.com.

Daphne Kis
Publisher