

# AORTA

## Voice, it's only natural

Voice enablement of “Net” applications and services is in a way the first phase of nirvana for pervasive computing. Back in 1999, when speaking to a WAP gathering in Barcelona, I mentioned that for WAP to succeed, it should really work with Voice. Some folks (especially from then Phone.com, now Openwave) stared at me like I was preaching some unknown cult. However, after my talk, several Europeans approached me to discuss what they were doing with voice-based access and why they agreed with me on the concept of multi-modal access. Since

then, WAP Forum, W3C, and VoiceXML forum have been working on a multi-modal framework, Openwave aligned with Conversa – a voice vendor to deliver integrated (voice and web) solutions and products, and in-general everyone agrees the importance of voice in the PVC world. Voice-based access will be more prevalent in personal-vehicle-ridden societies like the US and to some extent Europe. I just can't imagine the Japanese talking to their phones on a commuter train.

There are two schools of

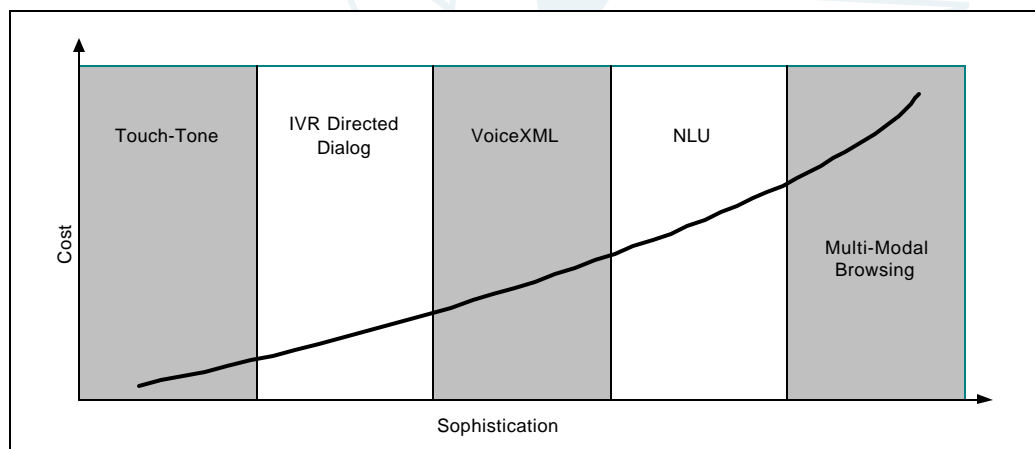
thoughts when it comes to Voice applications:

- a) Voice is just another channel in addition to wireless, broadband, etc.
- b) Voice applications are of paramount importance to corporations for their call-centers and VoiceXML is changing the way people design IVR systems.

Whatever may be the case, one thing is clear: **Voice solutions are here to stay.**

### Inside this issue:

|  |   |
|--|---|
| Voice—it's only natural                      | 1 |
| Waves of Technology—Expectations and Reality | 4 |
| Ask the expert                               | 6 |
| Ethermail                                    | 7 |
| Deconstructing 3G Mythology                  | 8 |

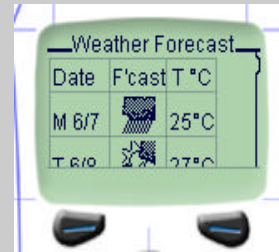


Cost vs. complexity of voice-based applications  
(Copyright: Sharma, Kunins, John Wiley & Sons, 2001)

VoiceXML represents the next generation in interactive-voice response applications (IVR). VoiceXML is to Voice applications what the Internet was to client-server computing. It allows us to be unbound by the constraints of the proprietary IVR programming languages and speech systems. Using an open architecture, one could mix-and-match best of breed solutions and not get locked down by “a” technology or vendor—you’ve heard the story before. In addition to VoiceXML, speech recognition accuracy has increased tremendously over the past couple of years, and of course with more processing power at hand, you could run more sophisticated algorithms quickly thus improving the chances of accurate recognition.

However, speech recognition by its very nature is an imperfect technology. What do I mean by this? When interacting with a computer or handheld, the way we interact with these devices basically remains the same for almost all the users. This is not the case with voice. There are large variations in languages, dialects, and accents. Our voice could vary if we are down with cold or are too tired. In addition, today, the conversations for most part have to be directed to have good recognition accuracy. If the conversation is free-flowing, computer has to not only identify the utterance but also the context (e.g. Could you *write* up a letter to Mr. *Wright* to give him the *right* directions). An average human would get the right writes to Mr. Wright but would a computer? If it’s **Natural Language Understanding-enabled**, it has a good chance. This can work in a demo environment which has had some neural-network training but to get it to the level where we can expect it to be part of most voice applications—is probably 12-18 months away. These advances will also tie in with the progress made on the multi-modal browsing front. It has the potential of increasing the usage of wireless applications tremendously. It would mean that we would also require speech recognition (SR) ON the device as opposed to SR on the server (which invariably is the case today). Most of the multi-modal applications don’t really need NLU since the interaction is directed but NLU can make multi-modal browsing so much powerful (start imaging yourself talking to devices) - think agents!

| Technology           | Computer-User Interaction (C: Computer, U: User)   |
|----------------------|--|
| Touch-Tone           | C: Press 1 for airline reservations  |
| IVR Directed Dialog  | C: Would you like to look for airline reservations or schedules?<br>U: Airline reservations.   |
| VoiceXML             | Voice-enabled “Net” Content  |
| NLU                  | C: United Airlines. How may I help you?<br>U: Could you please book me first-class from Seattle to Hong Kong? I want an early morning flight leaving day-after-tomorrow.<br>C: Done, Have a nice trip. |
| Multi-modal Browsing | U: Show me the weather forecast<br>C: Returns information on the phone   |



(Copyright: Sharma, Kunins, John Wiley & Sons, 2001)

Figure 1 (page 1) and Table 1 (above) provide a flavor of different generations of voice applications from touch-tone to Multi-modal browsing.

So, how big is the market? Any VC firm worth its money would fire back: is this another set of hyped technology? No doubt new technologies are always hyped, otherwise it won’t catch the attention of wall street and RVM (rear-view mirror) kids - isn’t it ironic? Kelsey group recently finished a study which seems to indicate that voice technology will be a \$16.3 billion market by 2005 (see page 5). Not bad, huh! (And that doesn’t even include the revenues made off transactions)

In this issue, you will also find a Q&A session (page 6) with a friend of mine, Jeff Kunins, Manager at TellMe Networks (considered to be *the* player in voice-based applications and services). Jeff is an accomplished spokesperson for the VoiceXML world. I am sure you will find his insights useful.

Internally, *Eswar “talk to me” Eluri* has been spearheading our voice efforts. His field services application is hosted through BeVocal. Try it out when you have a moment (see callout on next page for instructions) and give us your feedback. Can you start imagining scenarios and applications that some of our clients can start using or start planning right away? Can we build some targeted demos to help you make a point with the client? Let us know, we are all ears.

The WAP version of this application can be

accessed at <http://morse.luminant.com/AcmeFS/index.wml>. (developed by Hari Kanangi).

This application can also be accessed via Palm or any other handhelds having WinCE, PalmOS, EPOC, Linux, Neutrino operating systems. DB2 Everywhere components can be installed on the device, which can talk with the DB2 database on the server. It's a great example of the same application being accessed via multiple channels. The application server serves appropriate pages/content to the device requesting information. In future, we might extend the application to an iTV interface as well. You see the AORTA vision taking shape.

Getting back to voice, when is it appropriate to think of Voice-enabling applications and services, and more importantly when does it NOT make sense to think Voice.

I touched briefly on this in my last column. Let's discuss this a bit further. Speech applications can be used when:

- Voice is the most convenient mode of device input (example scenario: driving in a car)
- Navigation is complex and commands are embedded deep into the menu structure.
- Users have a physical disability

Speech applications aren't very useful when:

- The work environment is very noisy
- The user is in a situation of talking with devices and people at the same time (will lead to SR problems)
- It's easier to accomplish the task using other means of device input—keyboard, mouse etc.
- Content returned from the server is large and visually complex
- The task requires user to compare data items
- Information presented to the user is personal and confidential

We are fortunate to have Tod Knight, our esteemed CTO, pen an excellent two part series on **“Waves of Technology—Expectations and Reality”**. I am sure it will stir thoughts, ideas, and concepts in your mind. Please be sure to share them with us.

My sincere thanks to Tod, Jeff, and to all who contributed to this issue.

Your comments are always welcome.

Best wishes,

Chetan Sharma

Introduction to Voice  
Solutions—Class given by  
Eswar Eluri  
Tentatively scheduled for 25th  
April, 10:00AM PST  
Watch for more information.

## Field Services Demo (serving 100% VoiceXML)

- 1) Dial 1-877-33-VOCAL (1-877-338-6225)
- 2) Key in the access pin 0421
- 3) The user id is 6955716
- 4) This starts the field services demo. You will hear the welcome message and will give you some navigation commands which are always available.

### Commands:

- Help -> Plays the help on a particular item.  
Repeat -> Repeats a prompt  
Quiet -> Pauses a prompt by computer.

- 5) The Computer will prompt for User id.  
The userid and passwd is 123 and 123.  
You can either use the keypad on telephone or speak the entry, The voice input should be the digits 'one two three and a pause (not one hundred twenty three). Once you enter both, it will verify, and take you to the new prompt.

- 6) Choose a job type—Options: New, Open, All.  
Choose **All or Open**. If there are no new jobs, it will ask you to contact the dispatcher. You can hang-up or will be prompted if you want to continue

### 7) Job List

You will be prompted with the number of jobs in the list, followed by the commands available.

### Commands:

- Next -> Goes to the next job in the list.  
Previous -> Goes to the previous job in the list.  
Detail -> Goes to the details page and gives more information  
Start over -> Goes to the start of the list.  
Go back or Main -> Goes the Job types prompt

At this level, the Job ID, its site location and the status is available. For more details on the Job and to update the status, say Details at a particular job prompt.

### 8) Details section.

Lists all the details and asks user for an update on the status.

### Commands:

Accepted, Rejected, Completed, etc.

- Go back -> Goes back to the job list prompt  
Main -> Goes back to the job types prompt.

To exit at any time just hang up or say **"Bye"**

## Waves of Technology—Expectations and Reality (part 1 of 2)

- *Tod Knight, CTO*

In the movie *Pi: Faith in Chaos* a brilliant mathematician teeters on the brink of insanity as he searches for an elusive numerical code. For 10 years, he has been attempting to decode the numerical pattern beneath the ultimate system of ordered chaos - the stock market. When I originally read the description for this movie it reminded me of a similar analysis I have been doing recently around technology. While I don't **think** I'm teetering on the brink of insanity (yet), this could be attributed to a couple of things: those who are teetering are usually the last to know; I have not been maniacally focused on it for 10 years; or maybe I'm not brilliant(?) -- I'm certainly no mathemagician.

In part two of this series I will share with you some of the patterns I have uncovered while analyzing the technology market, but in this part I'd share some of the trends I have noticed and opinions I have formed.

It seems to me that with each new "Wave Of Technology" (WOT) -- from mainframe to client/server to Internet to wireless, and beyond -- that **the expectations for each "new wave" get higher relative to the realistic capabilities of the new technology.**

I support this assertion based on several common trends I have seen with each WOT:

**1. Things that were hard, if not impossible, to implement in the previous WOT are often orders-of-magnitude easier to accomplish in a new WOT. However, many things that were easy and even taken-for-granted are now either missing or even harder to implement in the new WOT. Quite often -- at least in the earlier stages -- this can offset any real net gain for a new WOT.**

For example, in the client/server WOT, giving "external users" access to systems was a HUGE undertaking. This involved creating and distributing disks with complicated setup programs; expensive support infrastructures (especially when you did not "own" those desktops); as well as large investments in banks of modems for users to dial into.

In the Internet WOT, we assume users have a standard (?) browser and you can simply point them at a

URL. However, in the early days, being able to quickly and easily develop these applications (relative to developing a similar application with client/server technology) was more complex; it was more difficult to make the applications scalable and reliable; and even today, "standard" browsers are less than consistent -- and don't even get me started on security...

Another example: with the advent of client/server technology the need for "hard copy reporting" and "batch processing" did not go away, but the capability to address these needs with early iterations of client/server technology were woefully inadequate. Note that these were ho-hum, bread-and-butter capabilities of the previous mainframe WOT.

**2. The "user interface" for a new WOT causes significant consternation, challenges, and learning curve, even as that interface is heralded as the next silver bullet, relative to the previous WOT.**

For those of you who have been around long enough to see the transitions, consider the challenges designing for a mainframe-modal-green-screen user interface versus a client/server graphical user interface (GUI). And then, almost as a step backward (or sideways?) the designing for an HTML-browser web interface. I believe we have yet another "side-step" in front of us as we begin to understand good interface design for the plethora of mobile devices out there, and the new ones yet to be seen.

While GUIs were heralded as being more productive than 3270 terminal screens, the initial iterations of GUIs looked awfully similar. The fact that "standards" were nowhere near in place for common GUI and non-modal look-and-feels meant that learning curves for new applications were even steeper. Then, just as users were becoming more accustomed to GUIs, the HTML browser interface (inside a GUI, Windows), meant yet another twist -- while browsers were more

standard, the way in which an application was implemented inside the browser was not. Now, we are trying to understand how to "shrink" our functional requirements into tiny screen real estate and a variety of

Have a cool idea, story to share, want to submit a piece or ask a question?

AORTA  
Content Deadlines:  
10th of each month

## (cont.) Waves of Technology—Expectations and Reality (part 1 of 2)

input methods.

### 3. Even before we fully address the security issues for the current WOT, the new WOT makes it orders-of-magnitude more complex.

Jupiter, a leading technology analyst and research firm, put out a report in November 2000 titled "Broadband & Wireless: Wireless Security" which discussed the major security flaws in today's wireless infrastructure. In this report they uncovered this key finding:

"Providers of mobile commerce services and financial institutions offering sensitive data over wireless devices must find solutions to security problems that exist in the mobile arena. *However, the availability of such security solutions in this immature industry is slowing the rollout of all mobile commerce services.*" [emphasis mine]

We saw similar issues as we migrated from mainframe to client/server. I would like very much to hear from anyone who has solved the "single sign-on" issue as it relates to their Internet/intranet/extranet systems, even as we are embarking on the wireless WOT...

### 4. The hype and expectations around a new WOT means that you either "wait until hardware/software vendors better support the features you need" (i.e. be "mainstream") or spend an inordinate amount of time pushing the technology to do things that will eventually be supported directly by vendors (i.e. be an "early adopter").

If you take the former route and wait, you may miss out on opportunities to realize at least some benefits sooner. However, if you take the latter route, you are often stuck with obsolete code and systems, or are forced to convert. To be truly successful, you must find that delicate balance between custom developing things you NEED TODAY, and deciding which things you can defer until they are natively provided by vendors. On this same topic, also review the previous point about security -- be careful how much of this is "roll your own" versus what you can get directly from vendors that is open and interoperable.

IMNSHO, for the wireless WOT, the "features we REALLY want and need" are going to be more hardware dependent than we have seen since the transition from mainframe to client/server. Think about it, the computers (laptops/desktops/servers) we use today are basically the same as

those we used for client/server, except that they are "bigger, faster and cheaper" -- except in the case of laptops, which are "smaller and faster".

I do not believe this will be the case with the wireless WOT. I believe most devices have gotten about as small as they reasonably can, and what we really need are entirely new form factors, not necessarily just smaller-and-more-converged versions of the same. We will need to see an order-of-magnitude change in the form factors and capabilities of hand-held devices to reap the most benefits from the wireless WOT. This means that we are faced with deciding how far we push the capabilities of the software available to us versus the capabilities that will be available to us in the (near?) future via the devices themselves.

### 5. New WOTs are often viewed or heralded as a replacement for the previous WOT.

Fortunately, this trend is less prevalent than in the glory days of client/server when it was claimed that mainframes would soon become extinct. However, some of this trend continues even today as people talk about no longer needing desktop computers and that all you will need is a PDA and/or mobile phone. In reality, new WOTs are typically extensions of the old, rely on some portions of the previous WOT, and will most likely need to co-exist in a practical fashion. The trick is to understand the right mix, and use the right tools for the job at hand.

Stay tuned for a discussion of the patterns that exist as new WOTs wash over us in the part 2 of this article. Perhaps we can uncover patterns in this chaos that will help us predict the future?

**Solutions HelpDesk:**  
You got questions, we got answers. Send e-mail to [Asksolutions@luminant.com](mailto:Asksolutions@luminant.com) for any solutions-related queries or ideas

**Ask the Expert** *(reprinted with permission)*  
- Jeff Kunins, Manager—TellMe Networks

**Q: Isn't VoiceXML still immature? Is it ready for "prime time" deployment in the enterprise?**

**A:** VoiceXML is a proven technology currently being used to deliver voice services such as 1-800-555-TELL, Shoptalk, Indicast, BeVocal and others. Ready for enterprise deployment, VoiceXML combines proven technologies in speech recognition, telephony, and Internet services. Companies such as IBM, Motorola, Lucent and AT&T have been honing the concept of simple declarative markup languages for speech applications for more than five years. Similarly, the Internet has achieved commonplace acceptance in the enterprise market and will process over \$30 billion in secure transactions this year. There is an overwhelming trend in enterprise deployments to more deeply embrace technologies such as XML, XSL, and HTTP as the universal transport. This paradigm allows companies to preserve flexibility and work more efficiently by cleanly separating data from the user interface. Developers build shared business logic once, then use standardized markup languages such as HTML (Web), VoiceXML (speech), and WML (WAP) to create the appropriate user interface for each device. VoiceXML is simply a commitment by technology leaders to adopt a universal open standard for Internet-powered speech applications (see Figure 1). Traditional Web services use technologies such as Perl, ASP, or Java Servlets to dynamically generate HTML by executing database and application logic on the server. VoiceXML brings this paradigm to the IVR market, giving developers an easy, integrated way to extend their services to the phone. VoiceXML applications are literally a new set of "pages" on a Web site that happen to describe a conversation rather than a visual interface. Companies can leverage VoiceXML to make voice an integrated component of their mobile strategy, authoring shared business logic once and investing new effort only in the specific user interface for each device they support.

**Q: Does VoiceXML support robust telephony applications or call center integration?**

**A:** The full range of call center applications and IVR tasks in deployment today can be built using a VoiceXML platform. For example, 1-800-555-TELL is a sophisticated voice application built entirely using VoiceXML. More advanced features, such as blind and bridged call transfers, outbound notifications, and intelligent call routing and agent "screen-pop" through integration with CTI middleware, can all be supported using a combination of VoiceXML and accompanying platform services also built using open Internet standards. The key lens through which to consider this issue is the analogy of the Web and HTML. HTML is explicitly designed to specify the visual presentation of an interactive application via the PC. Therefore, it contains appropriate constructs for tasks such as table layout, form input, and embedded images. It does not cover tasks such as database access, credit card processing, or personalization, which are typically handled by code running on the Web server. Products such as Microsoft Passport™ expose simple APIs to advanced shared services using open standards such as HTTP, cookies, and SSL. VoiceXML plays exactly the same role in the world of Internet-powered voice applications. VoiceXML has intrinsic constructs for tasks such as dialogue flow, grammars, call transfers, and embedding audio files. It even supports Voice over IP-based call transfers through the Session Initiation Protocol (SIP). Companies can build upon this foundation to provide additional shared services such as outbound notifications and call center integration through simple URL-based APIs analogous to Microsoft Passport™ or affiliate programs from Amazon and MapQuest.

*Jeff Kunins manages Tellme Studio, the premier open community and toolset for VoiceXML developers, and in addition helps define and articulate Tellme's strategic direction for meeting the needs of Enterprise customers. Previously during a four-year tenure at Microsoft, Jeff led a diverse set of platform and consumer efforts including Passport and Cinemania.*

## ethermail



Mediamaster from Nokia (on Linux OS) combines TV, Computer, and phone into one device.

- What does it mean to serve documents from a content repository onto a PDA/handheld device? Is this practical from a usability perspective?
- What sort of rollback/recovery mechanisms are appropriate for an enterprise web site, where the content that is served for a request is based on user preferences and personalization? What are the records retention and legality issues?
- If bandwidth is free, will there ever be conference room software that works for meetings?

Mike Alsup  
Leader, Content Management Solutions Practice

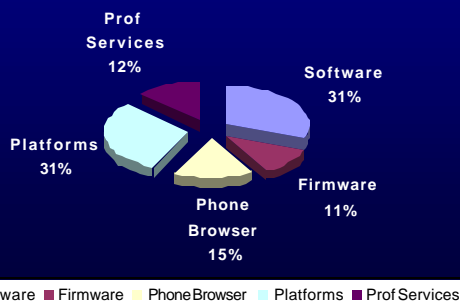
Mike

*You hit on three key factors in serving content to handheld devices— usability, personalization, and bandwidth. Till we get to flexible displays (where we can just roll down the display and connect it to a processor), we really can't serve content as we would do on a desktop. Ok, that was the obvious part. Assuming flexible displays don't arrive anytime soon for commercial markets, for all practical purposes, we are looking at Palm, PocketPC, Webpad kind of devices (things get worse for common use phones; though larger display phones are arriving, price will hinder mass-adoption). In terms of bandwidth (WAN), we can only expect 30-50 kbps using the best technology available (outside Japan). So, where does this leave us?*

*Three important things need to be considered: what can be reasonably served to the device. One involves **transcoding** the content to appropriate format and content type, device and network capability, and user preferences. The second involves **auto-summarization** of content — meaning shrinking the content into it's summary automatically. We can do the summarization part today by creating the content that way, but the auto part requires more work and more sophisticated technology. Specialized applications use that, but for this functionality to become an integral part of the server software, it will take some time. And, finally, I see **wireless LAN** technologies becoming more prevalent in the enterprise space, so instead of waiting for "true" 3G, we can use Wi-Fi (wireless fidelity, another name for IEEE 802.11) to get speeds up to 54MB/s though data rates deteriorate with distance. So, we would see more of webpad and tablet PCs like the ones shown on pages 7 and 8, having the power of a PC with the mobility of a phone.*

*In regard to verification of content delivery (essentially messages), one has to deploy a message queuing technology for delivery guarantees, so installing messaging packages (MQ Series, etc.) will be necessary.*

### Worldwide Voice Technologies Market \$16.3B by 2005

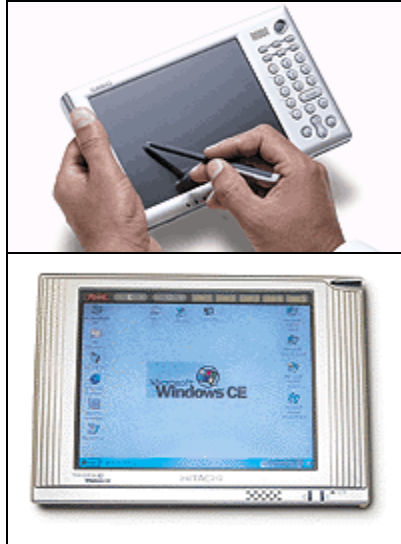


### Stat Focus

*Have an interesting stat or quote to share!,  
what are you waiting for?*

One should make sure to log key transactions (if not all) and link them back to the user and the device. This will help in non-repudiation.

NASA's advice on this is to keep logs for about 1 year, and then to conserve space, save them on a zip drive offsite. Using digital certificates, one can further reduce liability as they provides protection under law.



If bandwidth were free—oh! boy, wouldn't that lead to some shakeup. We are already progressing towards that with LAN infrastructure. We are moving towards a model where wireless is going to work much more closely with wireline to provide the maximum throughput. In as early as 1-2 years, most of the major inhabited spots in most advanced cities around the world (in places like Seattle, Sweden, London, Australia, it's already happening) will have 50MB/s or better bandwidth availability through a combination of wireless LAN, fixed wireless and fiber optic technologies, so if you are near an airport, mall, downtown, community centers, campuses, etc., you will have access to all the bandwidth you need. Won't NetMeeting do a much better job then?

Also, by then somebody will have designed the software that has conference room and peoples' calendars working together with the help of agents. So, for example, if 5 key people needed to be on the conference call, the bridge would automatically start bringing in people (connecting them from wherever they are using any device) - no need for wasting the first 15 minutes just trying to get people under the same bridge -- well, sort of.

Thanks,

Chetan Sharma

#### Book Recommendation

### Art of Innovation Tom Kelley

IDEO, the world's leading design firm, is the brain trust that's behind some of the more brilliant innovations of the past 20 years—from the Apple mouse, the Polaroid i-Zone instant camera, and the Palm V to the "fat" toothbrush for kids and a self-sealing water bottle for dirt bikers. IDEO's GM Tom Kelley presents a step-by-step approach to innovation.

## deconstructing 3G mythology

In the last issue, I promised to discuss the 3G market hype, especially the one promoted by the carriers. Why can't we believe what they are saying? Let's look at some *key points*.

### "We are rolling out first phase of 3G"

I first heard that phrase from an AT&T exec at a prospective client meeting where folks from IBM, Motorola, Symbol, AT&T, Elnk, and Luminant were invited to brainstorm wireless applications. When the person gave the timeline for such an installation, I didn't buy it. On further discussion, she was referring to GPRS (General Packet Radio Service) - a 2.5G technology with theoretical data rates up to 144kb/s (with practical limits). Sprint has started using the same language "initial phase of 3G is going to be rolled out later this year", tried to trick us, huh! It's 2.5G and NOT 3G.

### "It will deliver 2MB/s"

Any marketing rep in the wireless food chain will tell you: 3G == 2MB/s. Wow, you say. But it's a shared spectrum, meaning you can only have 2MB/s if you devote the entire capacity to one RF channel with one user hugging the cell tower. Does that mean higher bandwidths are out of question? No. I think it means that in order for carriers to come up with revenue-generating biz model, they will have to start paying attention to QoS (quality of service) -- they didn't have to deal with that stuff in the past. So, you have two variables: QoS and Bandwidth, and I think there will be pricing models based on the two. So, a user could have say, .95 QoS and 19.6kbps or .85QoS and 150kbps -- sort of bandwidth on demand -- but you pay for it. Along the same lines, we will not see more than 20kbps "effective data rate" initially from GPRS rollouts.

Download of a couple of average MP3 files is roughly equivalent to 3 hrs of voice call in terms of spectrum usage. This brings to another important aspect -- *battery life*. Do you really want to drain your batteries listening to MP3 files? Probably not. Not today.

### "We can build and they will come"

Business model for 3G applications and services is still unclear. Carriers are just hoping that somehow users will be ready to pay for applications that haven't been dreamt yet.

### "Underestimating the pain phase"

Geoff Varrall, one of the most highly respected industry analyst I have known, divides a wireless technology generation into three distinct phases of "*pain, pleasure, and perfection*" - each phase lasting for about 5 years --pain being, "oh! My god, the sky is falling", to perfection—"let's sit back, reap the rewards and move on to the next big thing". Almost every camp claims perfection in the pain phase. And history tells us that's almost never the case.