



Nuance Dragon dictation and search available for Apple iPhone from App Store

Large-vocabulary speech recognition for speech-to-text and search terms

Nuance Communications' Dragon NaturallySpeaking speech-to-text dictation software for the PC has been refined over the years and gets good marks from reviewers and users on its accuracy. In December, Nuance announced the availability of "Dragon Dictation" and "Dragon

Search" on the **Apple** iPhone, two separate applications available through the Apple App Store as a free download "for a limited time." The downloaded client software is supported by speech recognition within the network.

Continued on page 26

Microsoft's Bing for Mobile supports voice search

"Just say what you want and get it"

On December 2, **Microsoft** announced a "Bing Mobile" application for Windows phones, enhancing its Bing search engine. The application supports voice-enabled search using Microsoft's Tellme operation, as well as some location-based services. The basic paradigm, as Microsoft summarizes it, is "just say what you want and get it."

Bing Mobile, for example, lets a user speak to access maps, driving directions, and traffic

information. The new application also includes an improved auto-locate feature to show your current location, and makes it easy to speak to find a nearby business, a new restaurant, or a local movie theatre with the latest show times. The application handles complex spoken requests, such as "One Microsoft Way, Redmond 98052." One can choose walking or driving directions.

Continued on page 27

Vlingo is "everywhere"

New Blackberry version allows speaking rather than typing into any field

Vlingo has been one of the more successful small companies in mobile applications of speech recognition (see interview, p. 21). The company claims that its small-client, network-based speech recognition has been used over 100 million times by some four million users worldwide. The Vlingo application brings up a text box when a button is pressed on a supported mobile phone. One can then speak what one wants converted to text, with

keywords such as "email to Bill Meisel" initiating the dictation if it is to be a text or email message and other keywords to search the Web, update Facebook or Twitter status, and create a note to self, as well as speak a contact to call.

On December 10, Vlingo introduced Vlingo 4.0 Plus, featuring "Vlingo Everywhere" for the BlackBerry. With the new release, users can now

Continued on page 27

LumenVox and Digium offer \$7.99/month speech recognition with Asterisk IVR

Software-as-a-Service model

Telephone speech recognition licenses are typically sold with a per-port annual licensing fee, and one that is fairly steep. For users of the open-source Asterisk IVR software (a commercial version of which is sold by **Digium**), they can now pay \$7.99 per month for **Loquendo** speech recognition with no

long-term commitment (SSN, November 2009, p. 18). The \$7.99 price is equivalent to buying the normal LumenVox Lite license, which allows up to 500 words to be recognized at a time. Users who

Continued on page 28

Table of Contents

Nuance Dragon dictation and search available for Apple iPhone from App Store	1	Interact VoiceXML/CCXML supports high-density environments	11
<i>Large-vocabulary speech recognition for speech-to-text and search terms</i>	<i>1</i>	<i>IVR, voice portals, conferencing, voicemail, unified messaging, and prepaid services</i>	<i>11</i>
Microsoft's Bing for Mobile supports voice search	1	Vocollect voice-driven warehouse system delivers efficiencies	12
<i>"Just say what you want and get it"</i>	<i>1</i>	<i>Parts distributor, retailer, and construction equipment manufacturer benefit</i>	<i>12</i>
Vlingo is "everywhere"	1	Australian service provider Optus adopts SpinVox voicemail-to-text	13
<i>New Blackberry version allows speaking rather than typing into any field</i>	<i>1</i>	<i>"Spoken through SpinVox" tagline</i>	<i>13</i>
LumenVox and Digium offer \$7.99/month speech recognition with Asterisk IVR	1	Confirmit partners with Plum Voice to offer telephone-based surveys	13
<i>Software-as-a-Service model</i>	<i>1</i>	<i>Collect and report results in real-time</i>	<i>13</i>
Editor's Notes	4	Nuance healthcare business continues to expand	14
The "Mobile Internet": Technology driver for the next decade?	4	<i>Study shows importance of free-form content in medical reporting</i>	<i>14</i>
<i>Bill Meisel, Publisher & Editor</i>	<i>4</i>	Nu Echo's NuBot Hosted IVR testing platform available	17
EU-sponsored spoken language understanding project concluded	6	<i>Testing tools available free</i>	<i>17</i>
<i>LUNA project managed by Loquendo</i>	<i>6</i>	Nexidia expands speech analytics suite	18
Loquendo expands speech technology offerings	6	<i>DMG Consulting reports speech analytics market growth</i>	<i>18</i>
<i>Adds new application partners</i>	<i>6</i>	Sandata time and attendance product adds PerSay speaker authentication	19
Nuance highlights trends in automated customer service for 2010	7	<i>Bank Leumi rolls out password reset service with Persay authentication</i>	<i>19</i>
<i>Growth in proactive notifications, mobile customer care, and "cloud computing"</i>	<i>7</i>	Chinese call center systems integrator to offer Utopy speech analytics	19
Nuance technology behind AT&T voicemail-to-text service	8	<i>Utopy also announces a US customer, TantaComm</i>	<i>19</i>
<i>Option available for monthly charge</i>	<i>8</i>	TelStrat releases new version of its contact center solution	20
Promptu introduces voice dictation for mobile text messaging	9	<i>Incorporates phonetic-technology-based speech analytics</i>	<i>20</i>
<i>Includes full texting application</i>	<i>9</i>	J.D. Power and Associates 2009 auto navigation study	20
Telephonetics VIP creates £1million saving for hospital with outbound reminders	9	<i>Speech recognition improves satisfaction with system</i>	<i>20</i>
<i>Premises-based solution considered more cost-effective than hosted for UK hospital</i>	<i>9</i>	Ford demos of Sync at Best Buy prove successful	21
Verizon Business Open Hosted Speech Services available for IP telephony	9	<i>"Sync: Say the Word" ad campaign helped drive participation</i>	<i>21</i>
<i>Customers use Verizon's hosted speech technology, but host own application</i>	<i>9</i>	Interview with Mike Phillips, Vlingo	21
Parkmobile's mobile cashless parking uses VoltDelta's hosted platform	10	<i>"...the only limiting factor in what you can do on a small mobile device is the user interface"</i>	<i>21</i>
<i>Pay-by-phone mobile-phone service uses speech recognition</i>	<i>10</i>	VUI Visions	23
Voxeo partners with Remego and ICR	10	Continuous Automated Speech Tuning and the Return of Statistical Grammars	23
<i>Selling Voxeo contact center solutions in the Asia Pacific region and the UK</i>	<i>10</i>	<i>Roberto Pieraccini, SpeechCycle</i>	<i>23</i>
Aspect leverages internal capabilities in its customer support center	11	News briefs.....	28
<i>"Seamless Customer Service" and "Ask an Expert" services</i>	<i>11</i>		

Google adds Japanese speech recognition for mobile phones28

Google’s director of research Peter Norvig indicates Google will be using speech recognition in video search.....28

Google winning brand-image battle over Microsoft on mobile phones, according to research firm.....28

AT&T identifies coverage issues with iPhone app.....29

Runner’s app for iPhone from Abvio adds text-to-speech for Twitter29

Sabse Technologies enables Jingle Networks’ 1-800-Free411’s free international calls.....29

Syntellect announces a new forum for its developer community29

INX to provide Cisco voice portal solution30

SVOX adds Arabic text-to-speech30

Apple has a (well-hidden) text-to-speech option in its Preview software (delivered with the Snow Leopard OS).....30

HP, Real Soft, and USAN using GM Voices for prompt recording.....30

Frost & Sullivan comments on “behavioral biometrics,” including voice verification31

Nice Systems introduces multi-channel interaction analytics to extract insights from customer touch points beyond voice31

Market research firm predicts global automotive telematics shipments to be included in 84.6% of all cars by 2016.....31

3M introduces physician dictation on BlackBerry and Windows Mobile smartphones31

BigHand replaces WinScribe dictation solution at Australian law firm, WinScribe announces award for best legal dictation system32

Motorola introduces hands-free voice-directed mobility solution for the warehouse.....32

Grundig Business Systems dictation microphone features support for speech recognition software and low energy consumption32

Ginger Software uses text-to-speech as part of software to aid in writing32

Aircraft firm chooses Adacel for cockpit voice control system33

Barnes & Noble Nook e-book reader said to have hooks for future speech recognition33

Speaker identification plays role in Jakarta trial.....33

Google “translated search” allows finding information from sites written in other languages33

Lingo Media subsidiary Parlo to train China Southern Airlines’ cabin attendants in spoken English using speech recognition34

Toshiba develops a speech-to-speech translation system that can run entirely on a mobile phone.....34

Speech recognition techniques to check your heart? .34

TD Canada Trust and TD Waterhouse add web accessibility features.....34

Tigal offers voice control module for robots35

New draft of VoiceXML 3.0 spec published.....35

New working draft of W3C “Multimodal Architecture and Interfaces” available35

Call Genie to design, deploy and maintain an end-user facing directory assistance website for TDC Denmark35

Best wishes from Speech Strategy News for the New Year!35

Financial Notes 35

SpinVox takeover by Nuance rumored.....35

Nuance officers’ bonuses reported.....35

Call Genie announces closing of \$4.0 million convertible debenture financing36

Avaya completes acquisition of Nortel enterprise solutions.....36

For Further Information on Products Mentioned in this Issue 36

Google signs on as principal sponsor of the Mobile Voice Conference 2010 40
Introducing the decade of the Mobile Internet with a critical role for speech technology 40

Editor's Notes

The "Mobile Internet": Technology driver for the next decade?

Bill Meisel, Publisher & Editor

The investment firm **Morgan Stanley** issued a *Mobile Internet Report*, a 694-page report that in effect declared Mobile Internet Computing as the technology driver of the next decade, characterizing the 1990s as being driven by Desktop Internet Computing, the 1980s by Personal Computing, the 1970s by Mini Computing, and the 1960s by Mainframe Computing.¹ A preview of the report was presented on December 15 by one of the authors, Mary Meeker, a well-known analyst that made a similar presentation presaging the growth of the Internet.

This forecast is of course of interest to this newsletter, since it seems that mobile phones and mobile computing in general will increase use of speech interfaces. In addition, we are co-organizing the Mobile Voice Conference with AVIOS, to be held in April (www.mobilevoiceconference.com).

The significant work represented by the report deserves respect and makes some insightful points; however, it raises some questions. For example, a chart listing companies that are "well-positioned" to benefit from the Mobile Internet includes companies such as **Google** and **Apple**, and **Microsoft** is listed as a "potentially challenged" company. And why not note public companies that are relatively small today, such as **Nuance Communications**, which could benefit from speech technology being used on mobile phones. Previous computing revolutions have changed small companies to large companies (e.g., Google, Apple, and Microsoft).

Major themes in the Mobile Internet report

Let's outline a few points from the report:

- *Market impact of smartphones isn't fully measured by market penetration*; mobile Internet usage reflects the usability of the smartphone. For example, Morgan Stanley says the Apple iPhone and iPod Touch are responsible for 65% of mobile Internet usage, although they represent only 17% of global Smartphones.² This implies that usability is a key factor, but the long presentation makes no mention of speech recognition as a factor in Mobile Internet growth.
- *Material wealth creation and destruction should surpass earlier computing cycles*. The report notes that winners in each computing cycle often create more market capitalization than in past cycles, and that past winners often falter.
- *The Mobile Internet is ramping up faster than the Desktop Internet did*. Morgan Stanley believes more users may connect to the Internet via mobile devices than through desktop PCs within five years.
- *Five key factors provide the foundation of growth*: 3G adoption, social networking, video, VoIP, and "impressive mobile devices."
- *"Massive mobile data growth"* will drive the market. The focus of the report is on the data channels impact, as opposed to the voice channel.
- *In emerging markets*, mobile may be the primary means of access to the Internet.
- *Mobile phones are moving from a focus on voice communication to multi-purpose devices*. One chart in the report shows that the average American cell phone user spends 40 minutes a day on a mobile phone, making calls 70% of that time. The average iPhone user, by contrast, spends 60 minutes on the device but makes calls only 45% of the time. The rest of those 60 minutes are spent texting, e-mailing, listening to music, playing games, and surfing the Web.

Parsing the "Mobile Internet"

But what is the "Mobile Internet"? The report seems to emphasize the "Internet" in that phrase, treating the Mobile Internet as the Web accessed by a mobile device. I would emphasize the "Mobile" in the phrase as the key to growth. Our mobile device will almost certainly have a wireless connection, so it keeps us connected

¹ The full report is available from Morgan Stanley:

http://www.morganstanley.com/institutional/techresearch/mobile_internet_report122009.html

² Perhaps these numbers discount mobile email use as part of the Mobile Internet, since Research In Motion's Blackberry is particularly popular for this feature, and continues to show strong growth.

to others and to information sources; it will even tell us where we are and how to get somewhere else. We can take a mobile device everywhere, and, since it can always be with us, we can come to depend on it. (Most people have experienced the panic that rises when a mobile phone is misplaced or lost.) Since we have adapted to having our primary mobile device with us (almost certainly a mobile phone) almost all the time, that device will tend to increase the number of functions it provides, at the expense of other mobile devices such as audio players.³

To sound a theme familiar to readers of this newsletter, too much functionality on a small device strains its usability. In addition to the small form factor that makes the use of the Graphical User Interface and keyboards more difficult, there is the obvious need to free hands and eyes in many cases when mobile, with use while driving being the most obvious example. Even when not driving, the device is often held in one hand and controlled with the other, as opposed to the two hands used with PCs. Touch screens help, but, unless we evolve smaller fingers to adapt to the device, there are limitations to touch technology on small devices.

Options for using speech as a part of the mobile user interface continue to grow; see, for example, the recent expansion of offerings by **Microsoft** (p. 1), **Nuance** (p. 1), **Promptu** (p. 9), and **Vlingo** (p. 1 and interview, p. 21). An adequate user interface will allow the natural growth of mobile devices as Morgan Stanley anticipates, but it isn't the Internet per se that creates that growth. The Web is a well-established phenomenon of the last decade, as the report itself points out. It is mobility and making that mobility feasible that marks the new trend.

Limits to exponential change

Another conundrum raised by the report is the convenient description of one decade-long predominant trend in computing after another. Many areas of technology clearly grow exponentially, Moore's law of chip complexity being one of the more famous. The economist W. Brian Arthur, in his 2009 book *The Nature of Technology: What It Is and How It Evolves*, goes into great depth on how technologies evolve and why technology growth accelerates. (In part, it is because technologies are assembled from other technologies; and, as the toolkit of available technologies grows, invention becomes easier.) Yet, the apparent linear progress of computing breakthroughs belies this supposed acceleration. One could perhaps argue that the innovation of each decade requires some multiple of effort over the last to create, but that argument seems a bit contrived.

Perhaps the mystery lies within us—a technology must be used, and there is an interaction between the human user and the technology. Most humans I know don't change their habits exponentially; most are in fact a bit resistant to change. It takes exponential improvement in usability to persuade people to move (linearly?) toward adoption of a new technology. Automobiles may have evolved internally to be more reliable and fuel-efficient, but the driving controls for those automobiles—steering wheel, brakes, and accelerator—haven't changed fundamentally in decades.

Smartphones today rely on an understanding of the graphical user interface on PCs and the keypad on all mobile phones to make them acceptable to their owners. Most innovations are clever adaptations to a small device, rather than breakthroughs. As has been noted in previous editorial notes, the hundred thousand downloadable applications for the iPhone are symptoms of compensation for the failing of the basic user interface. If not, why don't we need 100,000 apps for the PC?

One could argue that the prediction of the importance of the Mobile Internet over the next decade requires that we overcome this resistance to change with a true and effective focus on innovation. The fact that all mobile phones have a microphone makes speech recognition a leading candidate to lead that innovation.

³ A potential hurdle is battery limitations, but I suspect this will be overcome in the long run by easily used induction chargers in coffee shops, in autos, and other places we frequent, chargers that don't require a physical connection.

EU-sponsored spoken language understanding project concluded

LUNA project managed by Loquendo

The LUNA project (Spoken Language Understanding in multilingual Communication Systems), a EU-sponsored initiative, recently concluded its three-year run. The project manager was **Loquendo** (see following article). Silvia Mosso of Loquendo, LUNA project coordinator, said, "We have successfully managed a complex and challenging project, working with leading research institutes and industrial partners in Europe. The results we have achieved demonstrate the excellence of the consortium and pave the way for new and more ambitious research topics." Prototypes used to test the approach were a tourism application in French, a help desk service in Italian, and a route planner for city public transport in Polish. The results are available for public use. A number of published research papers are listed on the LUNA web site (www.ist-luna.eu).

The LUNA project addressed the problem of understanding spontaneous speech in the context of automated telephone systems. The project aimed at developing robust technology capable of understanding customer's spoken requests spoken naturally, as opposed, for example, to highly constrained directed-dialog systems. The goal of the project was to address a new generation of spoken language interfaces, by proposing new methods,

algorithms, and tools to solve complex tasks and adapt to the context in which the dialog is taking place. A three-tier software architecture was used for Spoken Language Understanding in the project: (1) generation of semantic concept tags, (2) semantic composition, and (3) context-sensitive validation.

LUNA targeted an increase in the automation rate of telephone services. Mosso said, "The results we have achieved demonstrate the excellence of the consortium and pave the way for new and more ambitious research topics. We think that in the near future we will see an ever-increasing pervasiveness of speech interfaces, which will become widespread, not just in call center applications, but for accessing any information by voice."

Giuseppe Riccardi, Professor and Head of the Adaptive Multimodal Information and Interface Lab at the **University of Trento**, Italy, said, "In the LUNA project we have accomplished three important results. We have designed machine learning algorithms and set the state-of-the-art for robust spoken language understanding for multiple languages and different complexity applications. We have proposed potentially powerful semantic models for understanding language in context which could support next-generation problem-solving conversational machines."

Loquendo expands speech technology offerings

Adds new application partners

Loquendo, which provides speech recognition and text-to-speech (TTS) technologies, expanded its technology offerings in December. It also announced the use of its technology by additional partners in navigation systems and assistive technology.

MRCP server 64-bit Linux support

The Media Resource Control Protocol (MRCP) is a communication protocol supporting speech technologies on servers. Loquendo MRCP Server and Loquendo Speech Suite are server solutions for speech-enabling large-scale telephony applications, such as contact centers and message reading services. Loquendo released Loquendo MRCP Server 7.1.3 and Loquendo Speech Suite 7.0.8, extending support for 64-bit versions of Linux RedHat Enterprise 5.1, 5.2 & 5.3 and Linux CentOS 5.1, 5.2 & 5.3. Loquendo MRCP Server 7.1 is a backward-compatible evolution of Loquendo Speech Suite 7.0.

The Loquendo server-based products are also available for other operating systems, including versions of Windows. Loquendo MRCP Server and Loquendo Speech Suite are interoperable with the products of major IVR vendors, including Avaya, Nortel, Cisco, Genesys, Interactive Intelligence, Envoy, Siemens, Aculab, NMS, Voxpilot, App-line, Aizan, and others.

New speech recognition and TTS languages

Loquendo released its speech recognition software, Loquendo ASR, in Australian English—bringing the Loquendo Speech Recognition portfolio to 25 languages. Loquendo ASR is multilingual, speaker-independent speech recognition technology that offers several acoustic models specialized for the various application fields, such as telephony (fixed, mobile, VoIP) and noisy environments (in-car), as well as being available in compact versions

for embedded devices. Loquendo TTS for Australian was released in summer 2009.

Loquendo also released Loquendo TTS in Norwegian with the female voice of Vilde, along with a new Greek male voice and a new Australian female voice, increasing the company's TTS language portfolio to 28 languages and 68 voices. Vilde joins Grace, the new Australian English female voice, and Nikos, the new Greek male voice.

Loquendo TTS now supports the International Phonetic Alphabet (IPA), as well as X-SAMPA (Extended SAMPA). Transcriptions of text are automatically generated in IPA or X-SAMPA; alternatively, users can insert their own IPA or X-SAMPA transcriptions directly into their prompts.

Loquendo TTS to be used in Alpine navigation products

Alpine Electronics Inc., the manufacturer of automotive audio and navigation products, will add Loquendo TTS to future navigation products. The "embedded" version of Loquendo TTS will be used for reading driving directions, street addresses, and points of interest (POIs).

Loquendo TTS in DynaVox Mayer-Johnson assistive device

Loquendo TTS technology has been incorporated into **DynaVox Mayer-Johnson's** new Xpress communication device. DynaVox Mayer-Johnson provides communication and education solutions for individuals with speech, language, and learning disabilities. Xpress helps individuals living with the effects of stroke, autism, Down syndrome, traumatic brain injury, ALS, and apraxia of speech to communicate effectively.

The Xpress is small enough to fit in a pocket or purse. Twin speakers and Loquendo's synthetic speech delivered messages crisply and clearly in any environment. Loquendo's TTS allows Xpress users to add emotion to and punctuate their communication with laughter, crying, sighs, whispers, and other emotive indicators. The Xpress includes thousands of pre-programmed pages as well as allowing custom entries.

Loquendo TTS in Intel's speech-enabled handheld reading device

Loquendo announced that Loquendo TTS has been integrated into the **Intel Reader**, the recently launched mobile, handheld device which reads aloud printed text to assist those with dyslexia or other learning disabilities and those with vision problems. About the size of a paperback book, the Intel Reader converts printed text into digital text, and then reads it aloud using Loquendo TTS. The user can select reading speed, pitch, and choose from a selection of male and female voices. The device's menus are also enabled with Loquendo TTS for ease of navigation.

Louis Burns, Vice President and General Manager of Intel's Digital Health Group, said, "We are proud to offer the Intel Reader as a tool for people who have trouble reading standard print so they can more easily access the information many of us take for granted every day, such as reading a job offer letter or even the menu at a restaurant."

RoboBraille email service for the visually and reading impaired community

RoboBraille provides audio content to the visually and reading impaired community around the world. The company's e-mail service is based on Loquendo TTS. It transforms written documents into a variety of alternative formats for visually and reading impaired people: Contracted Braille, audio files in mp3 format, structured audio books in DAISY format and visual Braille. RoboBraille can also be used to convert documents in PDF, gif, and tiff into more accessible formats. RoboBraille is available free of charge to visually and reading impaired users in non-commercial contexts.

Originally a Danish service developed jointly by **Synscenter Refsnæs**, an institution in Regional Zealand, and **Sensus ApS**, a private company, RoboBraille is now available in Ireland, Cyprus, Italy, Portugal, and Great Britain in the framework of a pan-European consortium supported by the European Commission in order to verify its technical, commercial, and cultural viability. Currently, the email service supports seven different languages. Two more languages will be added in 2010.

Nuance highlights trends in automated customer service for 2010

Growth in proactive notifications, mobile customer care, and "cloud computing"

Speech Recognition Update spoke with three executives from the Enterprise division of Nuance Communications about trends in 2010 in their area:

Amy Livingstone, Senior Director, Marketing, for the division; Jeff Foley, senior marketing manager; and Aaron Kneiss, product manager, Nuance

Notification Hub. They highlighted three areas of growth:

- *Proactive notifications*: Outbound messages such as prescription refill notices and appointment reminders via mobile devices, e-mail, or voice messages. Kneiss noted that consumers regard these messages as a service, and they save companies money both by avoiding some inbound calls to agents and by reducing losses due to missed appointments and other problems avoided by the notifications.
- *Hosted IVR*: “Cloud computing” is a hot item in the press, and Nuance emphasized that hosted customer service offerings are part of that trend. Nuance On Demand is one such offering directly from Nuance, and Nuance technology is used by a number of other hosted/managed service offerings. Amy indicated that Nuance hosted services had grown during the recession, as companies saw the benefits of paying for usage and avoiding capital expenses. Kneiss noted that outbound solutions are often hosted.
- *Mobile Customer Care*: Nuance offers a non-speech option for mobile phones, providing mobile subscribers with tools to automatically resolve common problems directly on their mobile phones, such as making account inquiries, upgrading their accounts, and paying bills. Amy indicated that, currently, this built-in functionality on the phone only serves the mobile service providers, who can pre-load it on the phone, but it may eventually be an option for other companies. Kneiss noted that one company saw a 26% increase in call deflection using Mobile Care and another a 77% reduction for the company’s IVR system for covered tasks.

Beyond some of these developing trends in platforms and applications, Foley indicated that customer service operations are adopting more

flexible dialog approaches to avoid overly structured directed-dialog systems. In addition to minimally structured call routing technology that can use open-ended prompts (see VUI Visions, p. 23), there are intermediate options, he noted, citing Nuance’s “SmartListener” technology. SmartListener technology addresses out-of-grammar responses from callers. Nuance data indicates that many times callers provide answers that sound reasonable, such as “I’d like billing please,” but that are not in the precisely modeled list of acceptable responses known as the grammar. Out-of-grammar responses can cause problems such as endless repeating of the options or a request to repeat, an outright recognition error that causes an erroneous action, or frustration by the caller who perceives they have said something responsive to the question. Nuance has a large database of contact center interaction data that suggests out-of-grammar utterances outnumber other performance issues by a factor of up to 5 to 1. They have used this database to identify common out-of-grammar utterances and put them in the grammar. In addition, other techniques such as keyword recognition and “robust parsing” (which in essence ignore extraneous words before and after the content words) also form part of the SmartListener technology.

Livingstone noted that Nuance is also committed to its partners. These include companies such as **Avaya**, **Cisco**, and **Genesys** that use Nuance speech technology as part of their solutions.

An internal trend for Nuance is the integration of the acquired core **IBM** speech recognition technology into Nuance 10, the next release of the Nuance telephone speech recognition engine. Foley said that the IBM technology will be fully integrated and won’t be a separate product. As part of the agreement with IBM, IBM Research is aiding in the integration process.

Nuance technology behind AT&T voicemail-to-text service

Option available for monthly charge

AT&T recently launched a voicemail-to-text in its wireless plans for \$9.99 per month, allowing users to receive voicemail in English or Spanish transcribed to text (as text messages or email) and review messages in any order, as well as save them as text. **Nuance Communications** announced its Voicemail to Text platform is behind that service.

The company is leveraging the core technology in its Dragon NaturallySpeaking dictation product in a

number of areas for the telephone (p. 1). Michael Thompson, senior vice president, Nuance Mobile, said, “Our service was built with carriers like AT&T and their subscribers in mind—an incredibly reliable and scalable voicemail to text infrastructure, and a valuable and easy-to-use service for mobile consumers.”

Promptu introduces voice dictation for mobile text messaging

Includes full texting application

Promptu Systems Corporation offers multimodal talk-and-type mobile phone applications (SSN, February 2009, p. 14). Promptu purchases its core speech recognition technology from an unspecified vendor, and enhances it to tune it to accents and specific applications.

On December 23, the company announced ShoutOUT TXT, a dedicated text messaging application for iPhone with voice dictation, available for download in the App Store. Transcribed spoken messages are returned in seconds, the company indicated. Users can tap a word to see and select from likely alternatives before sending their messages on to the intended recipient. ShoutOUT

TXT features include a number of features to ease the use of text messaging in general, such as one-touch navigation to all recent incoming and outgoing messages.

The speech recognition is server-based and fully automated (no human agents). It is optimized for North American English and will improve over time with use.

There is a charge for spoken messages, but Giuseppe Staffaroni, Promptu's CEO, said that ShoutOUT TXT allows sending and receiving SMS messages at a cost "far lower than the standard rates charged by operators." A plan with 250 voice messages costs \$4.99.

Telephonetics VIP creates £1million saving for hospital with outbound reminders

Premises-based solution considered more cost-effective than hosted for UK hospital

Telephonetics VIP, part of **Telephonetics plc**, provides speech recognition and voice automation solutions for enterprises (SSN, February 2009, p. 12). Deployed across 20,000 telephony channels, Telephonetics VIP processes over 35 million minutes of speech recognition per annum on its carrier-class platform in both hosted and customer premises solutions.

In December, Telephonetics VIP announced that its solutions have helped Sheffield Teaching Hospitals NHS Foundation Trust to lower their "Did Not Attend" (DNA) rate. Patients make appointments but don't show up, leading to inefficiencies in the healthcare process and higher costs, as available professionals' time goes unused.

DNAs are a serious problem across the NHS (the UK's National Health System), costing an estimated £600m (an average cost of £101 per appointment at Sheffield Teaching Hospitals). The Trust considered a number of reminder solutions and chose a premises version of Remind+ from Telephonetics VIP. Sheffield Teaching Hospitals trialed Remind+ with 12 departments before a full implementation. The reduced DNA rate generated by the trial equated to a projected £460,000 annual saving for the trial departments.

The Trust considered Remind+ to be more cost-effective than hosted reminder solutions. Paul Harriman, Assistant Director, Chief Operating Officer Directorate at Sheffield Teaching Hospitals, commented: "Remind+ addressed our range of requirements to include group and individual reminders and was more cost-effective, as it has lower revenue costs and is not paid for on a per-reminder basis. This was a big advantage with the volume of reminders we planned."

Dave Amps, Director of Customer Solutions Architecture at Telephonetics VIP, noted that a critical advantage of Remind+ is the ability to gather information on the patient's intention to attend, re-book, or cancel the appointment. "This enables the Trust to rebook appointment times which become available again," he said.

Remind+ is being used by 60 departments to contact over 3000 patients each day. During a single six week campaign, the Trust used Remind+ to call 48,000 patients. 1,561 patients responded that they would not attend, allowing 1,561 appointment slots to be available for reallocation to other patients and saving over £157,000. This demonstrates that Remind+ is effective in reducing waiting lists as well as DNAs.

Verizon Business Open Hosted Speech Services available for IP telephony

Customers use Verizon's hosted speech technology, but host own application

Verizon Business, a unit of **Verizon Communications**, has long provided a range of managed services for customer service operations,

offering both hosted and premises-based options, with the flexibility to deliver customer calls to either traditional or IP-based contact centers (SSN,

September 2009, p. 8). One option is the company's Open Hosted Speech Services, which allow a customer to use Verizon's hosted speech engines and speech technology support software, but to manage and host its own application.

Until this month, Open Hosted Speech Services was available only for traditional TDM phone service. According to Tom Smith, manager, IP contact center and speech product management, Verizon Business, the company is now offering its Open Hosted Speech Services in an IP version. Verizon's Open Hosted Speech Services lets customers control and host their own speech applications, while they still gain the advantages of Verizon's speech platform. The option allows enterprises to better customize their speech services to interact with their customers without having to manage the speech services. (Verizon Business licenses Nuance speech technology.)

Smith noted that there are a number of advantages to hosted speech services. Verizon handles all the speech licenses and may be able to provide cost-effective options because of volume license purchases. Verizon handles all updates, and provides redundancy to increase uptime. Further, the hosted environment allows handling peaks (like the holiday season) without owning licenses that cover the peak usage. IP versions offer other benefits, including easier integration of multiple sources of information, e.g., other data within the enterprise to provide personalized services.

As many voice hosting services have reported, Verizon Business says its hosted operations grow during the recession, since the pay-for-usage model avoids capital expenses. Smith said, "The only thing worse than spending money right now is losing customers."

Parkmobile's mobile cashless parking uses VoltDelta's hosted platform

Pay-by-phone mobile-phone service uses speech recognition

Parkmobile is attempting to change the way we pay for parking around the world. The company's pay-by-phone solution (substituting a cell phone and credit card for pocket change to pay for parking meter use) has the potential to make life easier both for motorists struggling to find the right change and municipalities that want to reduce the cost of maintaining and servicing the meters. John Young, a Parkmobile executive in the UK, indicated that Parkmobile already has an 80% share in the Dutch market. In the US, cities such as Grand Rapids, Michigan, have adopted the solution. Service for parts of Canada is also planned.

Callers initially register for pay-by-phone by calling Parkmobile's toll free number and speaking their license plate number and credit card information. Once registered, callers simply dial in to indicate the zone they are parking in. The system immediately identifies the caller from their cell phone number making each transaction simple and easy. Motorists can also opt to receive text message alerts prior to the expiration time. The service covers all parking modes such as digital permits, cashless on and off- street parking, electronic enforcement, and

also supports seamless integration with congestion charging.

VoltDelta OnDemand (SSN, September 2009, p. 1) is providing the speech recognition and telephony services that enable the cashless parking service. VoltDelta currently handles more than 2.4 billion calls and 2 billion SMS messages per year and has the ability to manage peak call volumes. The on-demand model means that there are no capital expenditures required of any city deploying a Parkmobile solution.

Albert Bogaard, CEO of Parkmobile North America, Inc., said, "VoltDelta assists Parkmobile in delivering outstanding customer care for mobile callers in part by overcoming one of the most challenging issues in voice recognition; successfully recognizing the unconstrained string of letters and digits found on license plates."

Cities and municipalities struggle with costs associated with meter collections while also trying to maximize parking revenues. Parkmobile addresses both issues by automating revenue collection while making it more convenient to fill the meter on time.

Voxeo partners with Remego and ICR

Selling Voxeo contact center solutions in the Asia Pacific region and the UK

Voxeo Corporation provides compatible hosted and premises call center solutions, an approach it has

begun calling "Unlocked Communications" (SSN, November 2009, p. 8). In December, the company

announced two partnerships. One is a partnership and reseller agreement with **Remego**, a Singapore-based provider of IVR telecommunication and call center services, under which the two companies will work together to bring Voxeo's solutions to enterprise contact centers in the Asia Pacific (APAC) region. A second partnership is with UK-based **ICR**, a provider of contact center self-service applications, to deliver automated multi-channel customer interaction solutions.

Voxeo provides an open technology platform in IVR, instant messaging (IM), SIP, VoiceXML, CCXML, and mobile web support to create and manage unified communications and unified self-service applications. Remego will be delivering solutions that are platform-independent and multi-

channel. Remego has customers such as **Singapore Telecommunications** and **Great Eastern Life**. John Amein, senior VP of strategic partnerships at Voxeo, said that the companies have already successfully been winning business in APAC.

ICR specializes in speech recognition solutions and has delivered self-service IVR applications to many UK customers for a number of years. ICR's services range from initial business investigation, identifying where voice technologies can benefit most, through application development and platform provision. ICR has successfully completed Voxeo's product and service certification process. Andrew Bentley, Director at ICR, said that Voxeo's multi-channel support is a key advantage and "will change customer interaction in the future."

Aspect leverages internal capabilities in its customer support center

"Seamless Customer Service" and "Ask an Expert" services

Aspect's Unified communications (UC) applications for the contact center use software to target operational objectives with specific capabilities from the Aspect Unified IP and PerformanceEdge platform products. The applications deliver inbound routing, outbound dialing, voice portal and Internet contact, workforce management, performance management, campaign optimization, recording, and eLearning capabilities to help organizations enhance customer service, collections, and sales and telemarketing business processes. The company reported in December on its own use in its technical support center of two of these offerings: "Seamless Customer Service" and "Ask an Expert."

Ask an Expert uses the presence detection, instant messaging (IM) and conference calling capabilities of **Microsoft** Office Communications Server 2007 R2 to enable support engineers to use presence identification and skill criteria to find available experts in other departments to address customer questions. Seamless Customer Service together with Office Communications Server simplifies remote

support with desktop sharing, which increases collaboration to resolve issues more quickly. All voice mails are received through Microsoft Outlook email and can be easily forwarded and/or retrieved through mobile devices.

Aspect Technical Services is using Seamless Customer Service to utilize comprehensive inbound routing, voice self-service coordinated with live service, and Ask an Expert functionality to increase first call resolution and enhance the overall customer experience. With Seamless Customer Service, call hold times for Aspect Technical Services have been reduced by 76% to just 44 seconds per call. Additionally, since the initial introduction of the Ask an Expert capability, the overall first day resolution rate has increased by 7% and the time to reach a support technician with the correct skill set has improved by 8%. Teams using Ask an Expert to reach out to Aspect's experts across the organization have shown overall increases in overall customer satisfaction scores by 6%, measuring 4.79 on a scale of 5.

Interact VoiceXML/CCXML supports high-density environments

IVR, voice portals, conferencing, voicemail, unified messaging, and prepaid services

Interact Incorporated's VoiceXML/CCXML media platform, SPOT, provides high-density media processing with call control signaling, IP, and PSTN connectivity. This allows operators worldwide to deliver new voice and data service offerings, including Interactive Voice Response, voice portals,

conferencing services, voicemail, unified messaging platforms, and prepaid services.

SPOT was developed for use in environments that handle hundreds to thousands of voice channels. Standard testing for SoftSPOT, a pure software SIP-based VoIP Media Platform begins at a minimum of

500 channels for a server with dual 2.4 GHz processors. Interact claims that this is 9 times more channel support than the majority of solutions on the market and more than three times greater than Java-based systems. SPOT features linear scalability and no limit on the number of channels that are supported in a distributed environment. SPOT supports multiple telephony platforms such as **Aculab's** Prosody product line, **TelcoBridges** TB640 and T'media hardware, the **Diallogic** CSP, and is also available as a pure software, SIP-based platform. The company also offers supporting development tools: SPOTbuild, a graphical service creation environment; SPOTproof, a test automation and load testing tool; and an Integrated Development Editor, SPOTcheck.

The company recently announced a number of developments:

- **Voitra Technologies** provides live and automated agent software for next-generation VoIP networks. Voitra selected SPOT to provide a media server for their growing customer base for automated IVR applications such as operator services, voice dialing, wake-up calls, and flight tracking. Mark Stainken, President and CEO, Voitra, said, "It was imperative for our project that Voitra find a solution that offered not only scalability, but also complete redundancy, and

Interact's SPOT VoiceXML/CCXML based interpreter set provided this. Since implementing SPOT, we have seen some impressive numbers, and currently have a customer running 150 licenses on 1000 lines."

- Interact announced that it has entered into a global distribution agreement with **Conversif**, a company focusing on Telco-grade voice and video solutions for the Carrier and ISP markets. As part of the agreement, Conversif will become the first Value Added Reseller (VAR) to offer a full range of design, development, deployment, and support services for Interact's SPOT Platform.
- **ReceptionHQ** has deployed Interact's VoiceXML/CCXML Media platform SoftSPOT (the Linux-based pure software VoIP solution) to support their existing Computer Telephony Integration application in their answering services contact centers. Interact's SoftSPOT will power the services ReceptionHQ offers, such as Receptionist Plus, Message Center, and Virtual Office. ReceptionHQ helps companies create a Fortune 500 image by answering all calls personally and efficiently using the specific company name, then directing calls to the appropriate company employee regardless of where they are.

Vocollect voice-driven warehouse system delivers efficiencies

Parts distributor, retailer, and construction equipment manufacturer benefit

The products of **Vocollect, Inc.** highlight a market for speech technology for workers in warehouses that has repeatedly shown the value of the solution, with commensurate growth. Every day Vocollect helps over 250,000 workers worldwide to distribute more than \$2 billion dollars' worth of goods from distribution centers and warehouses to customer locations. The workers interact with warehouse management software by voice to pick orders and perform other tasks while their hands and eyes are otherwise occupied. A March 2009 report from **Datamonitor** assessed the 2008 global market at \$462 million.

One example is the parts center of **IHI Construction Machinery Limited**, a Japanese company that manufactures and markets large-scale construction equipment including mini-excavators, hydraulic shovels and cranes, and associated environment-related equipment. Vocollect Voice is used by IHI for cycle-counting, receiving inspection, storage, picking, and shipping inspection, supporting

parts control for approximately 60,000 items of varying sizes at its parts center in Yokohama. Before introducing the voice solution, the IHI parts control center used hand-held terminals or paper labels.

The company has achieved a 70% reduction in work errors from its one-year implementation of Vocollect Voice, helping the company attain a 99.993% operating accuracy. The company also realized a 46% average improvement in productivity, reducing the number of workers per shift by 50%. IHI is now studying the potential of using Vocollect Voice for other applications, including picking up orders on the product manufacturing floor and PDA input control of the results of the inspection of finished products.

In another example, **O'Reilly Automotive, Inc.**, one of the largest specialty retailers of automotive aftermarket parts, tools, supplies, equipment and accessories in the United States, announced that it is accelerating its transition to a paperless distribution center (DC) network by greatly expanding its use of

Vocollect voice systems. O'Reilly is currently using Vocollect Voice for picking in most of its existing DCs. All future new DCs will activate their operation with Vocollect by the end of 2010.

O'Reilly has 3,415 stores in 38 states and more than 44,000 employees. Greg Johnson, senior vice president, Distribution Operations, at O'Reilly, said, "Our productivity and accuracy levels continue to improve throughout our distribution operations with the use of voice from Vocollect. Last year we processed over 96 million lines using Vocollect Voice. Voice has helped us to reduce cost, improve customer satisfaction and overall business performance, as well as achieve a very positive return on investment."

In another example that shows the effectiveness of a voice solution in smaller storage areas, **Zetes** put Vocollect solution into the "back-of-store" operations in 130 **Argos** retail outlets. **Argos** is a large general-goods retailer in the UK with 750 stores. **Argos** is unusual amongst major retailers in the UK because its primary means of displaying

goods to customers is through a catalog. Customers browse through the **Argos** catalog, select items to purchase, pay for the items, and then collect the items from the in-store collection desk or have the item delivered to their home.

Since introducing the voice solution, **Argos** has seen significant process improvements through reduced time to the final sale point. Alan Jeffries, head of store operations at **Argos**, said, "Previously stores needed to put away products within each cage and then release the products for sale. Now individual items are released for sale as soon as they are put away – 'on shelf' literally means 'on sale.'"

Zetes is a large European integrator of voice solutions. The company has now completed over 800 voice implementations supporting an installed base of over 45,000 users. The in-store voice solution being rolled out by **Zetes** for use by the delivery assistants will comprise **Zetes'** 3iV application software and supporting management tools and Vocollect's Talkman T5 wearable computers, VoiceClient, and VoiceConsole software.

Australian service provider Optus adopts SpinVox voicemail-to-text

"Spoken through SpinVox" tagline

With rumors circulating that **SpinVox** might be acquired by **Nuance** (p. 1), the company continues to win adoptions of its hosted voice-to-text service, used most often in voicemail-to-text applications. In December, **Optus**, an Australian provider of communications services including mobile, local, national and long distance telephony, announced the launch of its Voice to Text service using **SpinVox**, which will allow **Optus'** mobile customers to convert spoken VoiceMail messages into text and receive them as a text message on their mobile phone. To distinguish from thumb-typed texts, every **SpinVox** converted message will appear within "speech marks" and carry the "spoken through **SpinVox**" tagline, ensuring customers recognize the difference in message style.

The caller's number is presented as the sender of the text to make it simple for customers to reply with a voice call or text. Jim Jaques, Marketing Director, **Optus** Consumer said, "**Optus'** Voice to Text service is the simple and convenient way to receive a message when you miss a call. Whether you're in a business meeting or you've got your hands full with children at the supermarket, sometimes you just can't answer the phone and instantly reading an important message on your mobile can make life much easier."

Anandh Maistry, Vice President Asia Pacific, **SpinVox** said, "Voice-to-text is a truly exciting category of voice service which is spreading across the globe at an impressive rate and is now the fastest-growing network service since SMS."

Confirmit partners with Plum Voice to offer telephone-based surveys

Collect and report results in real-time

Norway-based **Confirmit** provides worldwide SaaS services for surveys of customers, employees, and for market research. **Plum Voice**, which has both hosted and on-premise solutions supporting telephone speech applications, previously announced the availability of Plum Survey, a tool that allows

users to deploy a telephone survey using **Nuance** speech recognition (SSN, August 2009, p. 24).

In December, the two companies announced a partnership to offer a complete data collection platform to North American customers. **Confirmit** will add **Plum Voice's** IVR technology to complement its range of existing data

collection solutions that include Web, telephone, paper, and face-to-face. With Plum Voice's IVR technology, Confermit's customers can create a fully automated telephone-based survey that collects and reports data in real-time.

Gary Schwartz, senior vice president, marketing, Confermit, said, "By combining this crucial data

collection channel with Confermit's advanced reporting and alerting capabilities, companies can capture the Voice of the Customer immediately after an interaction – enabling them to take immediate action to retain customers and create positive customer experiences."

Nuance healthcare business continues to expand

Study shows importance of free-form content in medical reporting

Nuance Communications has a healthy business in healthcare, including a number of hosted and premise-based solutions for dictating medical reports, with the speech processed by speech recognition software (sometimes reviewed by medical editors and sometimes directly by the doctors). The company has clinical documentation and communication solutions in more than 5,000 healthcare organizations, and more than 250,000 physician users across the United States.

Nuance's healthcare portfolio reflects Nuance's apparent determination to supply complete, integrated solutions for healthcare. The portfolio, for example, includes closed-loop radiology management solutions to fully support a radiology practice, including radiology decision support (RadPort) to enable data driven, real-time e-Ordering, speech recognition reporting (PowerScribe and RadWhere), critical test result management (Veriphy) communication, and a business intelligence solution (RadCube) for utilization management, patient and outcomes analysis, as well as reporting clinical and operational trends. A number of Nuance solutions result from acquisitions. Nuance has clinical documentation and communication solutions in more than 5,000 healthcare organizations and more than 250,000 physician users across the United States.

Nuance and its partners reported a number of developments in its healthcare operations:

- The company announced the availability of Veriphy 4.0, a Web-based critical test result management (CTRM) solution for hospital diagnostic departments. The product does not incorporate speech recognition directly, but can of course manage reports created using speech recognition.
- Nuance joined the **Imaging e-Ordering Coalition**, an organization supporting healthcare providers, technology companies, and diagnostic imaging organizations. The Coalition works to promote Health Information Technology

decision-support (e-Ordering) as a solution to assure that all patients receive the most medically appropriate diagnostic imaging test for their specific condition.

- Nuance announced that **Renown Health**, northern Nevada's largest not-for-profit health network, and **UW Health**, which represents the five academic healthcare entities of the University of Wisconsin-Madison, have both added Nuance's diagnostic imaging solutions as part of their enterprise IT portfolio.
- **MedQuist**, a provider of technology-enabled clinical documentation services, announced that the radiology search tool from **Primordial Design, Inc.** has been integrated with SpeechQ for Radiology, MedQuist's real-time speech recognition system, which incorporates Nuance radiology speech recognition technology.
- MedQuist also announced that it has signed an agreement with Nuance under which MedQuist will now license Nuance's SpeechMagic speech recognition engine and processing software for use in SpeechQ of General Medicine, MedQuist's front-end speech recognition application for all medical disciplines.
- The **Sykehuset Telemark HF hospital** in Norway reported that it has been saving 900,000 euros a year on transcription costs since implementing Nuance SpeechMagic within its Electronic Medical Records system.
- **KLAS**, a research firm specializing in monitoring and reporting the performance of healthcare vendors, named Nuance's eScription on-demand platform for computer-aided medical transcription using background speech recognition, has received a 2009 Best in KLAS award, ranking #1 in the speech recognition category. eScription has achieved Best in KLAS six years in a row since 2004.

Nuance also sponsored a study suggesting that the current initiatives to provide digital healthcare records will face resistance from physicians if the

database systems are overly structured and don't allow the option of free-form comments summarizing the physician's observations and actions.

The study

Part of the reduction in medical costs sought by Congress and the federal government is targeted to result from improvements and expanded use of Electronic Health Records (EHRs). Nuance recently engaged more than 17,000 physicians in a survey to learn about physicians' hopes and concerns with respect to healthcare information technology, as well as to gauge their understanding of developing healthcare government policy, and nearly 1,000 responded. One key result highlighted the importance of capturing what the physician is thinking at the point-of-treatment, including a detailed physician narrative, and how this contributes to ongoing, high-quality patient care. Some results of the survey:

- **96 percent** of respondents voiced concern when asked how concerned they are "about losing the unique patient story with the transition to point-and-click (template-driven) EHRs";
- **94 percent** said that "including the physician narrative as part of patients' medical records" is "important" or "very important" to realizing and measuring improved patient outcomes; and
- **Less than 10 percent** of physician respondents said they were either "confident" or "very confident" that "the federal government's health information technology and reimbursement standards will lead to higher quality patient health records."

Over the last several months, the Centers for Medicare and Medicaid Services have been working closely with the Office of the National Coordinator for Health IT (ONCHIT) and the Department of Healthcare and Human Services (HHS) to define the incentive programs outlined in the HITECH Act, which will subsidize "meaningful use" of healthcare records. HHS was expected to publicly release a draft definition of electronic health record (EHR) "meaningful use" by the end of 2009. Nuance's survey found that there is concern that too much emphasis will be put on data capture and quantitative measures alone versus the capture of qualitative information that helps tell each patient's unique health story.

As part of the survey, respondents were also shown two versions of a de-identified patient's note (history of present illness, also known as the HPI Note), which was shared by Dr. R. Hal Baker, CMIO, Wellspan Health. The first note was created

by a doctor using speech recognition describing the patient encounter and care plan in narrative form. The second note (on the same patient, for the same visit) was created using an EHR point-and-click template, based on the structured elements selected by the doctor:

HPI Note #1 (dictated with Dragon Medical) – "The patient is a 74-year-old female who presents with a complaint of fall, 74-year-old female presents with complaint of neck pain, headache. She states that she had mechanical fall at home where she tripped and fell downstairs, approximately 9 steps and landed on her back. She complained of shortness of breath right after the event. She noted that she had pain in her left ankle and left knee. She is not sure whether she had loss of consciousness and the patient further complains of the pain in the right wrist."

HPI Note #2 (produced using an EHR template) – "The occurrence was one hour prior to arrival. The course of pain is constant. Location of pain: Head leg. Location of bleeding: None. Location of laceration: None. The degree of headache is mild. The other degree of pain is moderate. The degree of bleeding is negative. Mitigating factor is negative. Immobilization no backboard in place and no cervical collar in place. Fall description tripped. Intoxication: No alcohol intoxication. Location accident occurred was home."

When the surveyed physicians were asked which note they would "consider more valuable in treating this patient," 97% said HPI note #1. In addition, HPI note #1 was selected as the preferred note for addressing each of the following clinical communication objectives:

- "Driving high quality caregiver-to-caregiver communication," selected by 98%.
- "Recording the patient encounter, care recommendation and treatment history to safeguard them and/or their practice from medical/legal liability," selected by 93%.
- "Getting physician thoughts into the note – ensuring the medical decision-making is captured," selected by 97%.
- "Representing the uniqueness of the patient encounter – ensuring all relevant, personal information is captured and lives in the patient's health record," selected by 97%.
- Future visits with the patient "for understanding and recalling the patient's history," selected by 98%.
- 98% said HPI note #1 was "more complete and can be easily understood by the patient or another caregiver."

Peter Durlach, senior vice president, healthcare marketing and product strategy, Nuance concluded,

“The clear guidance from the physician community is that the Department of HHS should consider a requirement to ensure that each electronic patient note is not limited to templated text and structured data elements alone.”

Veriphy 4.0

Nuance’s Veriphy is an end-to-end healthcare enterprise solution for communicating critical test results from the lab, radiology, cardiology, and pathology departments to clinicians ordering the reports. Veriphy helps to ensure quick and auditable communication of patients’ urgent test results, allowing for reliable caregiver communication workflow, as well as expedited critical patient care. Veriphy is already helping nearly 175 hospitals nationwide to automate the delivery, verification, and documentation of critical test results. Veriphy 4.0 introduces more than 20 new and enhanced features, including several interface enhancements such as customizable drop-down menus, audio control for easy message recording, a screen minimization tool, and encrypted passwords.

Nevada healthcare organizations use Nuance radiology reporting

There is ample evidence that healthcare IT solutions that replace older technologies and manual processes can actually result in lower costs and improved care. One example of how technology can save the industry money vs. increasing spending is the shift from traditional medical transcription for clinical documentation to the use of speech recognition for radiology reporting.

Beyond the use of healthcare IT solutions for controlling cost growth, patient safety and compliance are also areas in which technology can contribute significant benefits. In 2006, nearly two-thirds of hospitals surveyed by **The Joint Commission** failed to meet the accreditation requirement for communication of critical test results.

Gary Wendt, MD, MBA, enterprise director of medical imaging at UW Health (which recently adopted a Nuance solution, as previously noted in this article), said, “Today, nearly 100 radiologists and 50 fellows and residents use speech recognition across multiple UW Health sites to report on more than 250,000 exams annually. Because we’ve added speech recognition to the radiology reporting workflow, we are able to create reports instantaneously, whereas five years ago it might have taken two-four days to finalize a report. Speech recognition has helped UW Health to significantly

bend the cost curve with estimated medical transcription savings at nearly \$400,000. In the next year, we expect to double the amount of radiology reports we create with speech recognition, and expect our savings to increase even more.”

Renown Health will deploy Nuance’s RadWhere, Veriphy, and RadCube solutions in March 2010. Christine Wells, Director of Outpatient Imaging, Renown Health, said, “Patient safety and compliance is core to our hospital’s mission. We believe that by adding the right mix of healthcare IT to our workflow, we will be able to improve the way we serve patients, create a safer environment for care, and enhance our ability to comply with industry regulatory requirements...As part of the roll-out, we will replace manual medical transcription, manual critical communication processes, and will gain intelligence on diagnostic imaging utilization and physician ordering habits; this knowledge will help us drive the best patient outcomes possible.”

MedQuist

MedQuist integration of a radiology search tool from Primordial Design with Nuance’s SpeechQ for Radiology will provide radiologists with a toolbar on their SpeechQ screen to enable on-demand, nearly instantaneous “Google-like” search of all patient reports at their site, eliminating the need to interrupt workflow to open other applications. The application can also be customized by the user to access archived images and reports from database systems as well as third-party radiology reference material and web sites.

Philip Zarboulas, Partner at Primordial Design, said, “The marriage of Primordial Search with MedQuist SpeechQ is a natural. Primordial Search takes advantage of SpeechQ’s structured reports to enable radiologists to search for keywords related to the specific clinical areas designated in reports, thus retrieving more accurate, clinically focused results. What’s more, because a site’s entire report archive is instantly accessible, radiologists can retrieve earlier images and reports more quickly and accurately.

Separately, MedQuist’s licensing of Nuance’s SpeechMagic speech recognition engine and processing software for use in SpeechQ of General Medicine, MedQuist’s front-end speech recognition application for all medical disciplines, is a natural extension of MedQuist capabilities. SpeechMagic currently powers MedQuist’s SpeechQ for Radiology application, used by thousands of radiologists to streamline the documentation process.

Based on its strong success in radiology and increasing customer demand for real-time interactive

speech recognition in many other hospital departments, MedQuist is launching a SpeechQ for General Medicine solution. This software application enables physicians in high-volume dictation specialties such as orthopedics, cardiology, surgery, and many others to realize the benefits of this technology. In addition, within an acute care setting, many report types such as Progress Notes, History &

Physicals, Procedure Notes, and Consultations lend themselves to interactive real-time speech recognition workflow. In real-time speech recognition, physicians dictate and self-complete reports in one sequence, bypassing the transcription process. The results include dramatic reductions in cost and turnaround time, while preserving physician preference for detailed narrative documentation.

Nu Echo's NuBot Hosted IVR testing platform available

Testing tools available free

Nu Echo develops speech recognition applications and tools for developing those applications (SSN, September 2009, p. 24). The company's NuBot Platform is based on a multi-tier software architecture (see figure below) integrating:

- The NuBot Integrated Testing Environment (ITE), a graphical client used to develop test scenarios, manage tests, and analyze results;
- The Robot Server, a middleware component that centralizes all call processing functions; and
- The Asterisk open-source telephony platform, which interacts with the application through a public or private network. (A low-cost monthly speech recognition option for the Asterisk platform is being offered by **LumenVox**, p. 1.)

The Robot Server communicates with the Asterisk platform through a standard API.

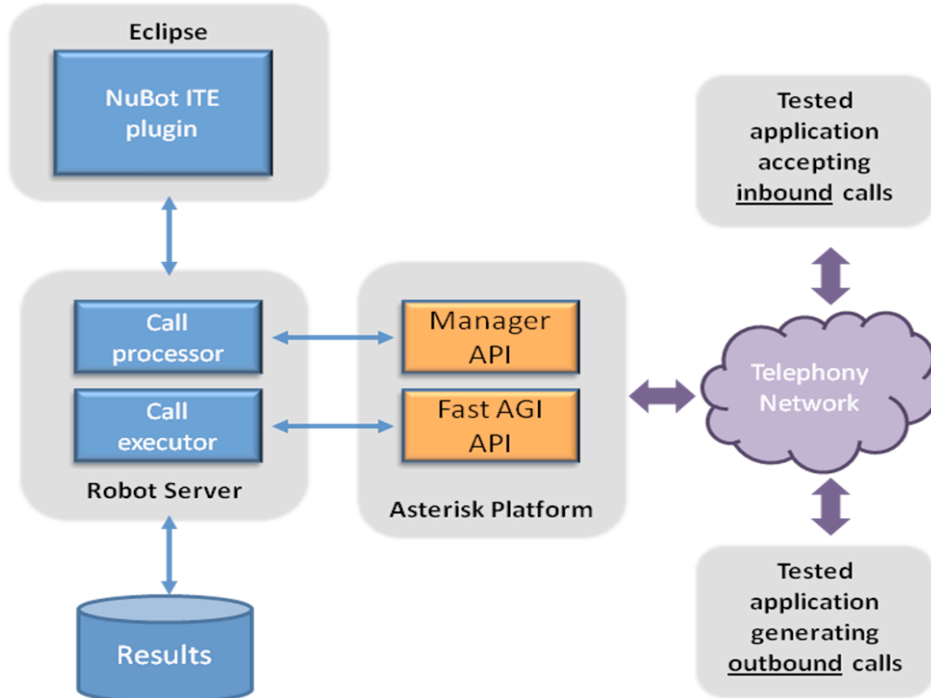
Nu Echo is now making the NuBot Hosted IVR Testing Platform generally available to the developer community at no cost at www.nuecho.com/nubot.

Yves Normandin, CEO of Nu Echo "We're convinced that making NuBot generally available will play a significant role in making more industry players realize the critical importance of automated testing in a rigorous IVR application development and deployment process."

With the NuBot Platform, developers can:

- Get a free copy of the NuBot Integrated Testing Environment (ITE), an Eclipse-based environment for developing test scripts of any complexity, managing tests, and performing extensive analysis of test results; and
- Run tests, using the NuBot Hosted Service, and only have to pay for their actual use of the service.

Another NuBot offering is the Test Authoring Model (TAM) to facilitate the development and maintenance of test scenarios. Users can create a



wide variety of scenarios quickly and easily. They can even reuse existing scenarios as building blocks

to create new ones, an important feature when trying to achieve the highest possible test coverage.

Nexidia expands speech analytics suite

DMG Consulting reports speech analytics market growth

DMG Consulting, a market research and consulting company, published its 2009-2010 Speech Analytics Market Report. DMG's research shows that the market grew from 25 commercial speech analytics implementations in 2004 to 1,764 at the end of 2008—a five-year compounded annual growth rate of 190%. DMG predicts that this market will continue to grow 45% in 2009, 40% in 2010, 42% in 2011, 32% in 2012, and 25% in 2013.

Nexidia provides speech analytics solutions using a phonetic approach. Jeff Schlueter, vice president of marketing at Nexidia, notes that the phonetic approach allows reviewing more interactions in call center recordings, and also supports adding levels of analytic capability that leverages the core recognition technology. Phonetic approaches pre-process the speech data to create a representation that is not “phonetic” in the classical sense, but includes a probabilistic representation at the phonetic level—a representation that can be searched more efficiently repeatedly for specific words and phrases than the raw speech data.

In December, Nexidia added additional capabilities to its Enterprise Speech Intelligence (ESI) software suite with Nexidia ESI-Quality. Schlueter said that ESI-Quality allowed contact centers to cost-effectively tap into 100% of recorded customer interactions to help measure and improve quality and agent performance. ESI-Quality is available immediately, and can be added as an option with any implementation of Nexidia ESI, including on-site deployments or as part of Nexidia's QuickStart and OnDemand hosted solutions.

Nexidia also announced a partnership with **Semafone**, a UK-based contact center payments security company, to address fraud and identity theft in the contact center environment. The huge volumes of credit card details being recorded and stored daily in contact centers internationally make it a prime target for fraud. The partnership between Nexidia and Semafone is designed to ensure that past and future credit card details are inaccessible by contact center staff.

ESI-Quality

The typical approach without speech technology is for a supervisor to sample a few of an agent's interactions, making the effectiveness of the review

questionable. Some systems use meta-data such as the length of the call or whether the agent indicates that the issue was resolved to make a more targeted selection, an approach that limits the evaluation to pre-determined criteria. Even with speech analytics systems that use word-spotting (rather than a phonetic approach), Schlueter claims that processing-time limitations require reducing the analysis to a small fraction of agent interactions.

The ESI-Quality process begins by defining “quality initiatives” in the Nexidia system. These are specific initiatives tied to performance objectives of the company, such as improving first call resolution (FCR) or reducing average handle time (AHT) for a certain category of call. Measurable performance goals are created to track these quality initiatives, with parameters set to show expectations of agent performance. ESI-Quality then automatically categorizes and measures all calls for each agent, and presents the information in a clear “Quality Portal” report that shows agent performance against the expectations for each quality initiative.

For example, a quality initiative can be established to track how well agents are managing average handle time for billing-related calls. Within the Quality Portal, supervisors can see at a glance how well their team is doing as a whole, and how well each individual agent is performing on this initiative. With a single click, the supervisor can bring up every call in the system that relates to that specific initiative for any agent to support coaching and performance improvement. Additional quality initiatives can be set up quickly as the business environment creates new corporate objectives.

“Nexidia's trademark has been our ability to provide scalable speech analytics solutions that extract relevant, business-oriented intelligence from customer interactions,” said John Willcutts, president and CEO of Nexidia. “With ESI—Quality, Nexidia is extending this intelligence down to the agent level, enabling contact centers to manage agent performance directly in line with corporate objectives.”

Nexidia and Semafone

With the introduction of new Payment Card Industry Data Security Standards (PCI DSS),

companies need new methods for dealing with the sensitive authentication data communicated during a transaction. Semafone's technical solution delivers secure voice transactions *during the live call* and ensures no new credit card details are stored. Semafone masks the DTMF digits entered by the

caller so they are never heard by the contact center agent nor recorded on the call recording system. Nexidia's audio search can find *historical* voice recordings of credit card numbers within an organization's recorded data so that they can be redacted.

Sandata time and attendance product adds PerSay speaker authentication

Bank Leumi rolls out password reset service with Persay authentication

Sandata Technologies, Inc. announced that it has selected Persay speaker authentication technology for its second-generation voice biometric engine for Santrax, a time-and-attendance product for field workers in the home healthcare and social services communities. PerSay also announced what they company cites as "the world's first eBanking web site that relies on Voice Biometrics for resetting customers' passwords."

Persay's VocalPassword is language- and accent-independent and integrates with existing VoiceXML platforms. It verifies a caller's identity in real-time using a spoken passphrase. The use of voice biometrics eliminates the need to give away personal information over the phone.

Time-and-attendance tool

In field work such as home health care, a worker will typically call from a site that they visit regularly, and the ANI of the phone can be used to verify from where they are calling. In order to avoid someone placing the call for the worker from that location, the caller's voice is compared to an earlier enrollment, insuring that the worker is at the location.

Santrax is available as a standalone product or as a component embedded in Sandata's comprehensive home care management software nHome and Integricare, its jurisdictional view dashboard for governments, managed care organizations, and other payers. Santrax enables home care agencies to

automatically verify that their staff is in the right place at the right time and capture tasks being performed, ensuring agencies are compliant with the prescribed plan of care and that patients are receiving the highest quality of care.

Bank Leumi password reset service

The Bank Leumi solution enhances the security of web sites and web-based transactions through multi-factor authentication. While current authentication processes mostly rely on information provided by customers ("something you know") that can be stolen or guessed, the new process combines biometrics ("something you are").

The self-service application enables customers to reset their eBanking password 24/7. It also replaces a costly and inconvenient manual questioning routine and eliminates the need to wait for a live agent.

Ariel Freidenberg, EVP Global Sales & Business Development at PerSay, commented, "The 'color of my cat' and 'my wife's maiden name' are questions that belong to the past. Looking forward, we see more and more financial institutions and other service providers adopting voice biometrics as an essential element in their security and risk management procedures. We expect regulation to drive multi-factor authentication for web, phone, and mobile transactions and believe voice biometrics is the optimal authentication factor answering this need."

Chinese call center systems integrator to offer Utopy speech analytics

Utopy also announces a US customer, TantaComm

Utopy, Inc. provides speech analytics for customer intelligence and performance optimization. The company announced a strategic partnership with **Grandsys**, a systems integrator for call center services and solutions in Greater China—a value-added reseller with an established customer base in recording, quality monitoring, screen analytics, and workforce management. Grandsys can offer Utopy's speech analytics platform, SpeechMiner, which

comes with prebuilt acoustic models and knowledge libraries for processing and analyzing human-to-human conversations in Mandarin and Cantonese.

Founded in 1992, Grandsys provides call recording, quality optimization, and workforce management solutions tailored for the Chinese market. The company combines expertise in contact center management and operations with extensive industry experience in banking, insurance, and

communications. Bill Yang, CEO at Grandsys, said, "Our customers are always looking for deeper, more comprehensive insight into their customers' expectations, employees' performance, and organizational processes. Together with Utopy, we are now able to offer our customers a comprehensive suite of innovative call center solutions."

Wisconsin-based TantaComm provides high-performance recording solutions for quality assurance, call verification, and regulatory

compliance. The company announced a partnership with Utopy to offer its speech analytics solution.

James Coan, TantaComm's Chief Operating Officer, said, "Speech analytics is an important component of TantaComm's 5-year business plan. When TantaComm evaluated whether or not to develop an internal solution, we ultimately decided that partnering with Utopy provided the highest value to our customer base...The SpeechMiner product is a layered enhancement to TantaComm's overall product set."

TelStrat releases new version of its contact center solution

Incorporates phonetic-technology-based speech analytics

TelStrat, a global supplier of contact center solutions and business call recording products, announced shipment of Engage 3.2, the latest version of its contact center solution suite. Incorporating TelStrat's previously announced phonetic-technology-based speech analytics product, Engage Analyze (SSN, May 2009, p. 22), the new release also brings new capabilities in other areas, including SIP support, on-demand recording, virtual server support, and a new low-cost option for **Avaya** VoIP recording.

Now being implemented at customer sites, Engage Analyze indexes and audio mines words and phrases buried in calls. Ed Templeman, director of marketing, TelStrat, said that the speech analytics solution is licensed from another company, which he declined to specifically name. TelStrat claimed that the phonetic approach to speech analytics makes the product an order of magnitude faster than competing

large-vocabulary speech-to-text systems. Engage Analyze is not dependent on finite dictionary and grammar models that require constant maintenance. This makes it easier to quickly search for new competitors, product names, slang, and other dynamically changing terms.

Avaya VoIP phone users can take advantage of Engage Record's new T-SPAN configuration. This configuration can eliminate some licensing costs. Engage 3.2 extends convenient, soft key-based, on-demand call recording to Avaya and **Cisco** VoIP phones. Ideal for schools, security, sales, and other quick-reaction environments, this lets call takers instantly access call recording directly from their phone, without requiring PC use. This makes it even easier for Avaya and Cisco users to access Engage Record's Conversation Save capability, which captures the total call from the beginning when activated at any time before the call's end.

J.D. Power and Associates 2009 auto navigation study

Speech recognition improves satisfaction with system

Real-time traffic and speech recognition are becoming increasingly popular features among factory-installed navigation system owners compared with 2008, according to the **J.D. Power and Associates** 2009 Navigation Usage and Satisfaction Study in December. Now in its 11th year, the study identifies six factors that contribute to overall customer satisfaction. In order of importance, they are: ease of use, system routing, system appearance, voice directions, navigation display screen, and speed of system.

Approximately 60% of owners report having speech recognition on their factory-installed navigation system. Among those owners who do not currently have the feature, two-thirds of respondents

express an interest in having it in the future. Overall navigation system satisfaction is positively impacted by the presence of speech recognition. Among owners who report having the feature, satisfaction averages 7.7 compared with an average of 7.2 among owners of systems without speech recognition technology.

The study finds that 26% of owners report having a factory-installed navigation system equipped with real-time traffic capabilities—twice the proportion of owners in 2008 who said the same. Real-time traffic capability has a considerable impact on overall navigation system satisfaction, as satisfaction averages 8.1 (on a 10-point scale) among owners

with the feature, compared with 7.3 among those without.

“Among the 10 navigation systems with the highest levels of customer satisfaction, all of them have real-time traffic capabilities, and a majority are equipped with voice recognition—underlining the positive effect these advanced features have on overall navigation system satisfaction,” said Mike Marshall, director of automotive emerging

technologies at J.D. Power and Associates. “With 39% of consumers in the study reporting that they would have purchased another vehicle if navigation wasn't available (up from 31% in 2006), there is a growing segment of automotive consumers who are driven by technology. Integrating and delivering advanced user-facing technology in vehicles will continue to gain importance among manufacturers fighting for market share.”

Ford demos of Sync at Best Buy prove successful

“Sync: Say the Word” ad campaign helped drive participation

Ford Motor Co. is expanding its partnership with **Best Buy** after a pilot program in Dallas resulted in increasing interest in Ford brands. Ford dealerships teamed up with 30 Best Buy locations to offer free demos outside a Best Buy store in a Sync-equipped vehicle. About 70% of those who participated in the demos over the past 30 days say they are now more likely to consider buying a Ford, Lincoln, or Mercury product. About 80% said that understanding how to use the hands-free Ford Sync system improved their overall opinion of Ford. Ford dealerships will now work with local Best Buy stores in Pennsylvania and California to hold customer clinics.

For Best Buy, the pilot acted as a traffic generator, with more than 1,600 consumers conducting mobile

phone upgrade checks over the 30-day program period. Prior to the owner clinics, 36% said they would consider a phone purchase from Best Buy Mobile. After the clinics, that number jumped to 72%.

In conjunction with the Texas pilot program, Ford launched its “Sync: Say the Word” ad campaign featuring robotic “voice agents” that act as personal assistants as voice commands are activated. The print and online campaign is aimed at educating buyers about the simplicity and power of the Sync system. The ads target young, tech-savvy drivers. The campaign's online component includes eleven instructional videos that highlight each Sync feature.

Interview with Mike Phillips, Vlingo

“...the only limiting factor in what you can do on a small mobile device is the user interface”

Mike Phillips, co-founder and CTO of Vlingo, was interviewed by Bill Meisel in late December. Mike has been active in the speech technology world for over twenty years. He started his career as a researcher first at Carnegie Mellon University and then at the Spoken Language Systems group at MIT working on core technology for automatic speech recognition. In 1994, he founded SpeechWorks based on technology that he and others had developed at MIT. Over the next ten years, Mike and team grew SpeechWorks from a small startup in a new market into the market leader in the now established market for speech-enabled call center solutions. SpeechWorks was responsible for many of key innovations in use today in the speech recognition systems deployed throughout the world. In 2003, SpeechWorks was acquired by ScanSoft (now named Nuance). Mike joined ScanSoft as CTO and oversaw technology integration and development across the product groups. In 2005, Mike left ScanSoft to spend a year as a visiting scientist at MIT before starting Vlingo in the summer of 2006.

Please outline Vlingo's core strategy.

Ever since we started the company, our goal has been to create a broad multimodal user interface for mobile devices. It's our view that as mobile devices and networks become more and more powerful, the only limiting factor in what you can do on a small mobile device is the user interface, and that speech has an important role to play. But, we also realize that speech is not the only interface people are going to want to use so we have not just been trying to create speech interfaces, but rather adding speech functionality to existing user interface modalities on the devices we work with.

The other part of this goal—to make it a broad interface—meant that we had to get away from the notion of speech-enabled applications, and instead think of speech as a user interface on the device that should be able

to work with any application. This led us to do things like get rid of application-specific grammars, and instead adopt a strategy of broad Hierarchical Language Models and aggressive use of adaptation to allow us to adapt to whatever people speak to any application.

Our applications deployed so far have not fully achieved this vision—mainly because we have been constrained by what we can layer onto existing mobile user interfaces and existing mobile operating systems. But, we believe that we and the rest of the market will be continuing to move in this direction of a broad multimodal interface across mobile devices.

What is the business model? How will Vlingo ultimately generate profits?

Vlingo is pursuing a number of avenues for revenue generation. In our consumer applications (available either by downloading our application from mobile application stores or in some cases our application is preloaded on phones), we have a combination of free and paid products. We have over 3 million users of these products so far, and an increasing number of end-users choosing to pay for the premium features. We also think that the functionality we provide will become part of the phones and services, so are also pursuing relationships with carriers, handset makers (most notably, we signed a recent deal with Nokia), and application developers.

As you note, Vlingo's application was pre-loaded on some Nokia phones (SSN, October 2009, p. 1), a step that could be interpreted as making it part of the basic user interface of the phone. Do you see that as a trend (or at least a goal)?

So far, the cases where we have been preloaded have been the Vlingo application as a third-party application running on the existing phone operating system (Symbian Series 60 in the case of the Nokia preloads). But, yes, as mentioned above, we do think that over time Vlingo functionality should be more tightly integrated with the overall phone platform. This is necessary to make speech part of this broad user experience—not only does the speech functionality need to be integrated in a way that a user can speak into any application, but if it is truly part of the operating system, then application designers can start to take into account the fact that users can speak to their applications and may make some different design decisions to better optimize their applications for this use case.

Vlingo announced a "licensing agreement and strategic alliance" to incorporate AT&T's Watson speech recognition technology into the Vlingo network-based service (SSN, October 2009, p. 1), estimating at that time that the conversion would be completed in Q1 2010. Is that estimate still valid, and what impact will it have on Vlingo once completed?

Yes, we are nearing the completion of our integration of AT&T Watson speech technology in our server-side infrastructure and will be deploying our first commercial products using Watson in Q1 2010. We are doing a lot of work with the AT&T Labs researchers and are seeing some very nice accuracy and functionality improvements over our existing technology. In addition to accuracy improvements, the end users will start to see new functionality in our products based on the technology we are licensing from AT&T.

Vlingo's basic model is to allow dictation into a text box, taking that a step further with your recently announced ability to dictate on Blackberry phones into any text box (p. 1). Do you plan to add any interactive speech dialog as the interface evolves?

Our new "Vlingo Everywhere" functionality on Blackberry phones is a good example of what we mean by making the speech functionality broadly available across applications. It turns out that the Blackberry operating system gives us enough hooks to allow speech input into any text field. So, with our latest product release on Blackberry phones, users can freely mix speaking, typing, and editing in any application—without the application developer having to do anything. (Although there are some limitations based on how applications interact with the operating system, and there are a few applications where this doesn't work. We thought of calling the feature "Vlingo Wherever We Can" but the marketing team didn't like that.)

This does give us broad speech-enablement, but doesn't provide the sort of interactive speech dialog that you are asking about. Unfortunately, this is hard to do in a broad way. We do however include top-level routing. So, if you say something like "send message to Bill, thank you for including us in your newsletter," we will route that to an SMS function which will pop-up a screen to let the user complete the task (including

speaking “send”) to send the message. We are planning on increasing the interactivity of this sort of core functionality—especially for the case where people are using their phone while driving.

How adaptive to the user is your technology? Can you take full advantage of the mobile phone being in effect a “personal telephone”?

In order to achieve high accuracy, Vlingo makes use of significant amounts of automatic adaptation. In addition to adapting the HLMs (Hierarchical Language Models) that eliminate the need to use constrained grammars, the system adapts to many user and application attributes. Vlingo learns the speech patterns of individuals and groups of users, new words, which words are more likely to be spoken into a particular application or by a particular user, and pronunciations of words based on usage. Adaptation is also applied to individual users (for example, the system learns over time that a particular user tends to ask for pizza) as well as across users (a first-time user with a southern accent benefits from other users who have spoken into the system with a southern accent). Unlike other speech recognition technologies that require intensive manual effort to tune recognition inputs, Vlingo adaptation is automated and comprehensive, leading to continual improvements for users.

Any final comments?

We’ve made a lot of progress over the past few years, but feel that we are still just at the beginning of what is possible. The key next step is to work more closely with the mobile phone makers and carriers to include speech earlier in the design process of future mobile interfaces and operating systems. Given the recent success of mobile speech applications by us and by others, we are seeing that the phone makers and carriers are getting increasingly interested in this, so we expect to see these more integrated approaches in the market over the next year or two.

VUI Visions

Continuous Automated Speech Tuning and the Return of Statistical Grammars

Roberto Pieraccini, SpeechCycle

*In this guest column, we ask designers skilled in creating Voice User Interfaces to highlight a particular aspect of VUI design inspired by actual deployments. In this issue, Roberto Pieraccini, Chief Technology Officer, **SpeechCycle** (SSN, December 2009, p. 1), suggests that, even in the most directed dialog interactions, a well-trained Statistical Language Model will easily outperform a rule-based grammar by allowing more varied responses, and discusses issues and solutions in using SLMs more widely. Roberto has been involved in spoken dialog technology for more than 25 years, both in research as well as in the development of commercial applications. Prior to joining SpeechCycle, Roberto was the manager of the Conversational Interaction Technology department at the **IBM Thomas J. Watson Research Center**. Before that, he led the Natural Dialog R&D group at **SpeechWorks International** (now Nuance). Earlier, he joined the Speech Research Group at **AT&T Bell Labs** and later **AT&T Shannon Labs**. Roberto began his career as a speech scientist with **CSELT**, the research center of the then Italian operating telephone company, after completing his doctorate in engineering from the **Universita degli Studi di Pisa**, Italy.*

Statistical grammars, commonly dubbed “SLMs” (for Statistical Language Models) by IVR practitioners, have been known to the speech research world for almost 30 years. However SLMs started to make their first steps into the IVR world only relatively recently. On the other hand, rule-based grammars, often written in an XML dialect known as SRGS (Speech Recognition Grammar Specification) have been used since the first attempts to deploy speech recognition in the early 1990s. Rule-based grammars are king for directed dialog, and only for situations where an open prompt⁴ needs to be played, are SLMs brought into the equation with a lot of effort and mystique. But what the IVR world often ignores is that, even in the most directed dialog interactions, say plain YES/NO questions, a well-trained SLM will easily outperform a rule-based grammar, all other things being equal.

⁴ AT&T’s “How may I help you?” is the classic example of such an open prompt.

So, why are we using handcrafted, rule-based, XML grammars at all if we know that SLMs would work better? There are several reasons for that. First, building an SLM is not as easy as writing rules in XML. You need data, and it is the kind of data that you don't have when you first build a new spoken dialog system. And even if you had data—I mean recorded utterances as responses to each prompt—you would need it transcribed. On top of that you would need to provide for each transcribed training utterance—in a process called *annotation*—a semantic tag: its meaning. So while instructing transcribers to do the right thing can be straightforward, teaching annotators the correct utterance-tag mappings may be very challenging. And assuming you have your transcribed and annotated utterances, and you have cleaned all of the transcriptions and annotations to transform them into a consistent set of training data—and you need lots of them—now you have to build an SLM. How do you build an SLM?

Building an SLM requires you to have a special set of programs that do two things. The first is creating a properly called⁵ *statistical language model*, in other words a way to tell the recognizer which sequences of words are legal. While rule-based grammars explicitly list all the legal sequences of words, a statistical language model does it in a ...well...statistical sense [1]. This is accomplished by computing what people of the trade call *n-grams*, which are the probabilities for any possible word—at least for all the words that appear in the training utterances—to be preceded by any possible *n-1* long sequence of words. So, if $n=3$, as it typically is in practice, the statistical language model computes, for each word, the probability of being preceded by all the possible sequences of 2 words. Thus, if you had 1,000 words in your vocabulary—and typically you need more than that for an open prompt, less than that for a directed dialog prompt—the program that computes the statistical language model has to compute 1000 (all the possible words) times 1000x1000 (all the possible pairs of words) probabilities, in other words 1 billion probabilities! Don't worry ... you may not be able to find examples of all the possible 1 billion triplets of words of your measly 1,000-word vocabulary even if you search the whole Web. But the statistical language model training program has to provide a number, the estimate of a probability, even for the most unlikely triples, like *yes maybe computer*⁶. But again, don't worry. There are programs that do that for you, and you can buy those programs; you can get them for free from some open-source packages published on the Web, or if you are versatile enough and not afraid of a little math and some algorithms, you can build them yourself. But of course, even if you had the best of programs for building the best statistical language model, you have to fiddle with a number of parameters in order to get the best out of it. And that may not be easy. But that's not all.

Building an SLM is not just about constraining the recognizer on all the possible sequences of words. In a spoken dialog system you don't need just the words that were spoken by the caller, but a semantic tag, a symbolic output from a set of *slots* that has a meaning for the call-flow at the particular prompt. For instance, if the prompt is asking a simple YES/NO question—for example “Have you paid your most recent bill?”—you want the recognizer to return either a YES or a NO. If the caller says *yes*, you want the recognizer to return YES; if the caller says “you bet” you want the recognizer to return YES; if the caller says “no way” you want the recognizer to return NO, and so on. If your training utterances are semantically annotated, on top of being transcribed, SLM builders create what is called a *semantic classifier* [2], in other words a program that takes as input the string of words recognized by the speech recognizer, and returns one out of a number of slot identifiers. Semantic classifiers are built from large samples of transcribed and annotated utterances. Again, you can buy one of these programs, or you can get it from some open-source projects, or if you feel adventurous in some non-trivial math and some non-trivial algorithms then you can build it yourself. And even if you buy it, you still have to adjust parameters and do some non-trivial tuning if you want to get the best performance.

In short, this lack of data and expertise has made using of SLMs unpopular especially for directed dialog solutions since the early days of speech IVR technology. Think about how much easier building an XML rule-based grammar is in comparison.

⁵ Although the *statistical language model* is only a part of an SLM, the industry term SLM (Statistical Language Model) indicates the full ability to decode the meaning out of free-form utterances, or *natural language* utterance, typically the responses to an open prompt.

⁶ In fact, when an uncommon triplet of words is not found in the training set, statistical language models approximate that probability using some heuristic considerations. There should not be any triplet of words with a zero probability, since that would a-priori exclude that triplet, however uncommon, for being ever recognized.

But there is another reason for SLM's lack of popularity. If you create a rule-based grammar, you can see what you did. You can immediately understand why the recognizer did not recognize the phrase "maybe yes" spoken by an undecided caller—perhaps because that phrase was not in the grammar—and you can promptly make the necessary modifications in a matter of minutes. You cannot easily do that with statistical grammars. N-grams, probabilities, and statistical classifiers are inscrutable at first sight. If something goes wrong, you need some understanding of the statistical machine learning theory behind the SLM in order to fix it. Also, the common notion—common and widely accepted in research as a logical and experimental fact—that SLMs always outperform rule-based grammars if trained on the right data, is not wholeheartedly embraced by IVR practitioners. There is confusion between the performance of an SLM in an open prompt situation, and the performance of a rule-based grammar in directed dialog. Of course the latter works better than the former, but only assuming callers always say what is in grammar, which is not always true. But this notion muddles the notion that a properly trained SLM will outperform a rule-based grammar in a directed dialog situation, especially when callers say things that are out of grammar. Yes, a well-trained SLM will have at least the same performance, and most likely outperform a corresponding rule-based grammar. So, why are we using rule-based grammars at all? As the previous paragraphs elucidate, it has nothing to do with grammar performance and everything to do with the high price of admission that SLMs entail.

And here comes the idea. What if we could provide a way to create and tune statistical grammars *automatically*, and use them for every context in a dialog, either open-prompt or directed, in place of traditionally handcrafted rule-based grammars, and do that continuously, while the application is deployed? After all, except for transcription and annotation, there is nothing that strictly requires the continuous labor of machine learning and speech scientists to build SLMs, while that is not always true for rule-based grammars. Yes, speech scientists run experiments to determine the best set of parameters, but they can create programs that run the experiments for them and decide which best selection of parameters to pick. Yes, they condition and clean the data, making sure that there are no inconsistencies in the transcriptions and annotations. But they, the machine learning and statistic speech experts, can create programs that do that for them. They can also create programs that help reduce the cost of human transcription and annotation by automating it when possible. Experienced speech scientists who are also computer scientists and machine learning experts can work on programs that tune speech grammars, rather than working on speech grammars themselves. And programs, unlike humans, can handle an abundance of data, and the speech grammars can get better and better.

What I described is the concept of automated tuning [3], or *grammar factory*, which is a service that can take a constant flow of log data from your IVR and give you, continuously, better and better grammars. How much better? Well, that depends on other factors, such as the design of the call flow, the prompts, the task, and other things that can give rise to poor grammar performance if not properly done. But while the grammar factory is looking at improving grammars, it can also flag behavior that would require the inspection of expert VUI designers and speech scientists. Will the grammars keep improving indefinitely? Certainly not. After a while, after enough data has been processed, their performance will settle on the maximum possible performance, which may not be 100% accuracy because of all of the other factors that can influence speech performance. But one thing is sure. It would be very hard, if not impossible, to reach that theoretical maximum performance using old-fashioned, handcrafted grammars tuned by hand. And if any change occurs in the application—a new prompt, a new strategy, new products, a new language—the grammar factory will, automatically and relentlessly, adjust for that and guarantee, in a short time, the attainment of the best performance for the speech recognizer. This is a new step ahead in the direction of machines that truly understand speech and continuously learn from what they hear.

References

- [1] Young, S., "Talking to Machines (Statistically Speaking)," Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002), September 16-20, 2002, Denver, Colorado.
- [2] Evanini, K., Suendermann, D., Pieraccini, R., "Call Classification for Automated Troubleshooting on Large Corpora," Proceedings of the 2007 IEEE ASRU Workshop, Kyoto, Japan, December 9-13, 2007
- [3] Suendermann, D., Evanini, K., Liscombe, J., Hunter, P., Dayanidhi, K., Pieraccini, R., "From Rule-Based to Statistical Grammars: Continuous Improvement of Large-Scale Spoken Dialog Systems," Proceedings of the 2009 IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP 2009), Taipei, Taiwan, April 19-24, 2009

Nuance dictation for iPhone (cont.)

Continued from page 1

The dictation application allows speaking free-form text for emails, voice notes, text messages—any form of text, since the iPhone software allows easy cut-and-paste from the clipboard to any other application, including Facebook and Twitter. The separate voice search function doesn't expect grammatical text, and searches multiple sources with one request, a feature that would be useful even without the speech functionality.

The announcement isn't just one more voice application for smartphones. Nuance already has a major presence in the mobile phone and voice control market with simple applications such as voice dialing and predictive text, as well as more ambitious solutions such as voice support for navigation systems such as the **Ford Sync** (p. 21). Nuance speech technology is most likely behind the voice control that is part of the iPhone, although all Nuance will say officially is "Nuance licenses technology to Apple for a variety of their products." The new applications take speech recognition to a higher level of flexibility, and the core functionality is very likely to be expanded over time.

The acoustic and language models have been optimized for the telephone application. Michael Thompson, senior vice president and general manager, Nuance Mobile, said that both applications can be used immediately, with no enrollment; an enrollment of one's voice is recommended for best

accuracy on the PC version of Dragon. Both applications do adapt to the user's word usage (the Statistical Language Model) and voice/accent (the acoustic models) as they are used. Accuracy will thus increase with usage.

Dragon Dictation

The figure below shows the process for initiating dictation and how the results look. A "stop" button will cause the application to stop listening.

Thompson said that the response was "incredibly fast," within one or two seconds, particularly if the channel is WiFi. Thompson indicated that the software is optimized for the **Apple** iPhone and the iPhone channel.

The PC correction interface has been modified for the mobile phone. To correct an error in the transcription, the user clicks on a paragraph, the words light up, and alternative interpretations can be viewed on a drop-down menu. Edits are used to improve later speech recognition.

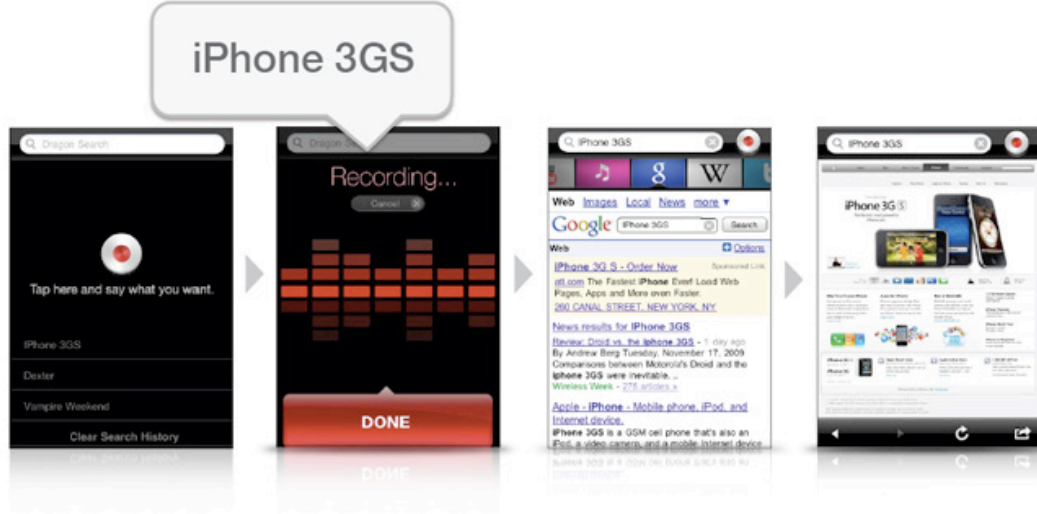
Among the adaptive features is the downloading of the user's contact list so that names can be more accurately recognized. This has caused a bit of controversy re security, but Nuance has pointed out that only names, not numbers, are downloaded, and are stored safely. In response to concerns, Nuance made downloading of contact names optional rather than automatic.

Thompson said that Nuance planned to sell the application through its usual channels. The company supplies many software options to handset manufacturers, including Apple.



Dragon Search

The Dragon Search App is illustrated in the graphic below. The voice-enabled search app for iPhone allows users to simply speak a search query and in seconds the “Dragon Search Carousel” displays results from a variety of sources: Google, Yahoo!, Bing, iTunes, Twitter, Wiki, or YouTube. The system does a simultaneous search of all sources, Thompson indicated, and one can see the results from each by clicking on the appropriate icon.



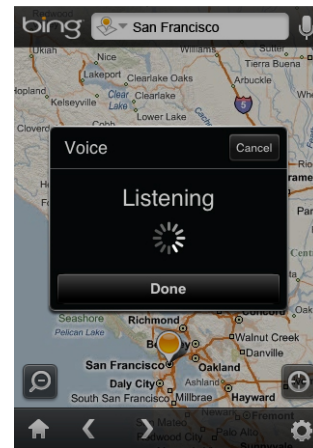
For example, one can say “Coldplay” and the Search Carousel will direct users to buy Coldplay songs from the iTunes Store, check out Tweets about Coldplay, watch a Coldplay video on YouTube, check the band’s history on Wikipedia, or even find Coldplay tickets on Google.

The speech recognition accuracy is enhanced by what Thompson called “dynamic updating.” Search terms that may be common because of breaking news are added to the vocabulary automatically.

Microsoft’s Bing for Mobile (cont.)

Continued from page 1

The new Windows Mobile Bing search client can be downloaded at discoverbing.com/mobile on a PC or m.bing.com on a Windows Mobile phone. But Microsoft isn’t limiting its ambitions to Microsoft-based phones. The company also released a version of Bing with voice search for **Apple’s** iPhone, available free in the iPhone App Store. A version for the BlackBerry is also available.



Vlingo (cont.)

Continued from page 1

...speak text into any field where their BlackBerry already allows them to type. The new release allows entering text by speaking within virtually every downloaded and native BlackBerry application, and even into text-boxes on websites. The company is currently developing Vlingo for other mobile devices and operating systems. Dave Grannan, president and CEO of Vlingo, noted, “The versatility and safety that Vlingo has added to smartphones by allowing

users to speak to their devices is what has made it one of the top application downloads for BlackBerry users.”

The basic Vlingo functionality is free. Vlingo Plus can be downloaded for a limited time for a discounted price of \$14.99 (later \$19.99). Current 3.0 Plus users can upgrade to add Vlingo Everywhere for a discounted price of \$9.99.

According to a Vlingo spokesperson, early use data for test versions shows that the new Vlingo Everywhere feature accounted for 42% of Vlingo total usage. Instant messaging is the most popular

application category for Vlingo Everywhere, accounting for 14% of overall usage. BlackBerry Messenger, a form of text messaging, accounted for 11% of overall usage, followed by speaking into web

browser pages at 6%. Other popular applications include calendar entry, address-book entry, Facebook, and Windows Live Messenger.

LumenVox and Digium (cont.)

Continued from page 1
need more words can subscribe to a LumenVox Full license, which allows unlimited words, for \$19.99 per month. Both prices include access to all of the features offered by the LumenVox Speech Engine, as well as support for its six different languages and dialects.

The LumenVox Speech Engine has been available on the Asterisk telephony platform for more than three years, and the Software-as-a-Service (SaaS) version uses the same Asterisk interface written by Digium. LumenVox President Edward Miller said the decision to offer the Speech Engine as a

subscription service reflects the growing demand for SaaS offerings in the marketplace. Miller added, "Because we have reduced the cost of entry to an unprecedented low price, anyone who has ever been curious about speech recognition can feel free to give it a try."

The usage-based model may be particularly advantageous for uses that are periodic. For instance, IVRs that perform outbound political polls see heavy volume immediately before elections that drops off after the election. Miller noted that these sorts of applications would be well served by no-commitment subscription licensing.

News briefs

Google adds Japanese speech recognition for mobile phones

Google announced at an event at the Computer History Museum in Mountain View in December, California, a Japanese-language voice search product that allows Japanese speakers to say a query into a cell phone. The product was previously available in English and Mandarin Chinese.

Google's director of research Peter Norvig indicates Google will be using speech recognition in video search

The siliconrepublic.com web site interviewed Google's director of research Peter Norvig when he was in Ireland to give the prestigious Boole Lecture at the University of Cork. Norvig co-wrote with Stuart Russell *Artificial Intelligence: A Modern Approach*. In response to a question, he said that Google will do more with searching video, including searching the spoken content of the video.

In response to a question on the future shape of human-to-computer interaction, he mentioned that speech recognition and computer vision are important. With respect to mobile devices, Norvig recognized that the small screen is a challenge and mentioned speech recognition as an option.

Army & Air Force Exchange Service to use Voxware warehouse software

Voxware is a supplier of software for voice-driven warehousing operations. In December, the company announced that the Army & Air Force Exchange Service (AAFES) has awarded a five-year contract to Voxware for the deployment of its Voxware 3 software product. Voxware was selected after a competitive evaluation process. The solution operates on **Intermec** CN3 voice-enabled wearable computers. To date the Voxware 3 software product has been deployed in 3 of the 11 AAFES distribution centers worldwide.

Google winning brand-image battle over Microsoft on mobile phones, according to research firm

In a report on "Facebook and Google Most Desired Brands for Mobile," **Strategy Analytics** investigated interest in the ability to use various applications from mobile devices. Survey respondents were more interested in having access to **Google** text search compared to **Microsoft's** Bing text search. **Facebook** ranked much higher than social networking rivals **MySpace** and **Twitter**. "Personalized content and services [such as **Yahoo!**'s offerings] are more desirable than access to media brands [such as TV network sites],"

commented Paul Brown, Senior Analyst in the Strategy Analytics User Experience Practice. “Yahoo! rated highly in both the USA and UK, as did websites **MapQuest** and **Weather.com** in the US, and **MSN** in the UK.” Christopher Dodge, User Experience Analyst at Strategy Analytics, added, “In terms of television brands, consumers predominantly want access to the same entertainment channels that they watch at home (e.g. Fox, BBC1), rather than news channels and sports dedicated to their mobile devices.”

“The key finding is that Facebook and Google are the most desired brands to have on a mobile phone in both the U.S. and Britain,” said Chris Schreiner, senior analyst of user experience practice at Strategy Analytics Boston office. “In both regions Facebook (first in the UK) and Google (first in the US) were ranked as the top two brands respondents were interested in having on their mobile phone,” he said. “The results for Google hint at the popularity that Google’s Android platform can achieve among smartphone owners, and the need users have for mobile search.”

Ecuador firm voice portal for international calling integrated by Interlancompu, VoiceXML browser from I6NET and speech recognition from Verbio

Corporación Nacional de Telecomunicaciones del Ecuador (CNT), an Ecuadorian national fixed-line carrier, has implemented a 116 phone self-service solution for all its international calls. The a voice portal is an IVR application using speech recognition to select the country, city, local number. The solution was integrated by **Interlancompu** using the VXI* VoiceXML browser from **I6NET**. The voice user interface was designed using Eclipse VoiceXML generation tools. **Verbio Technology** provided Spanish TTS and speech recognition engines.

AT&T identifies coverage issues with iPhone app

AT&T recently caused talk that its data network was getting over-burdened by iPhone use in New York by stopping sales of the device over a weekend. After this led to speculation over poor performance in the canyons of New York streets, the company resumed sales and refused to detail the reasons for the freeze. According to an earlier statement by AT&T, however, an independent research firm’s driving tests showed that AT&T had reduced its 3G dropped-call rate by 12% over the past year.

The company is using its customers to find out where there are demands for greater capacity. With the free AT&T Mark the Spot app, customers can note where and when a problem occurred. AT&T will use the metrics to direct its network-building dollars.

AT&T recently dropped a lawsuit against rival **Verizon Wireless**, which portrayed AT&T’s 3G wireless coverage areas in a series of new ads that AT&T called “misleading.” Perhaps it reconsidered when the CEO of Verizon said that “the truth hurts.”

Runner’s app for iPhone from Abvio adds text-to-speech for Twitter

iPhone fitness application developer **Abvio** announced Runmeter 2.0, a running application for the iPhone 3G/3GS. Runmeter turns an iPhone into a GPS stopwatch. Runmeter 2.0 also offers built-in Twitter text-to-speech, so runners can hear Twitter replies from their coaches, friends, and family during training. Runmeter 2.0 is available now from the Apple iTunes Store for \$4.99.

Sabse Technologies enables Jingle Networks’ 1-800-Free411’s free international calls

Jingle Networks operates the ad-supported directory assistance service 1-800-FREE411 (see interview, SSN, February 2009, p. 18). In December, **Sabse Technologies Inc.**, a VoIP, telephony-as-a-service (TAAS) company, announced that Jingle Networks has selected Sabse’s TAAS platform and global termination network to enable the free five-minute international calls offered to 1-800-FREE411 users.

By simply saying “free call,” callers to the toll-free number can also make a free five-minute call to almost anywhere in the world. Jingle Networks CEO Scott Kliger, said that the integration and rollout of international calling as a feature of our 1-800-FREE411 service has been smooth and worry-free. “That allows us to focus on our primary business of growing and serving our base of users and advertisers,” he noted.

Syntellect announces a new forum for its developer community

Syntellect announced the availability of the Syntellect CT ADE developer community forum (<http://forums.syntellect.com/>). CT ADE is a widely used communications development environment (SSN, September 2009, p. 12). “Many members of our CT ADE community are already working together, helping

each other leverage the power and flexibility of the CT ADE platform along with their developed applications. We wanted a tool to nurture this interaction,” said Holger Stoltze, Syntellect’s senior product manager.

Syntellect CT ADE is a rapid application development (RAD) toolkit that provides a set of program building blocks that are easier to use and faster to learn than traditional APIs. CT ADE offers complete access to **Dialogic** telephony hardware/software interfaces and control of speech products.

INX to provide Cisco voice portal solution

INX Inc. announced that it has been recognized by Cisco Systems as a Cisco Unified Customer Voice Portal (CVP) Advanced Technology Provider. CVP provides self-service and agent-assisted services to a company’s customers. CVP supports touchtone or speech recognition.

INX provides unified communications and data center virtualization solutions for enterprise organizations. The company’s services are centered on the design, implementation, and support of network infrastructure, including routing and switching, wireless, security, unified communications, and data center solutions such as network, storage, and server virtualization. Customers include enterprises, as well as federal, state and local governmental agencies.

SVOX adds Arabic text-to-speech

SVOX, which offers a variety embedded speech solutions (SSN, December 2009, p. 5), announced the immediate availability of the Arabic language across the range of SVOX Automotive and SVOX Mobile speech text-to-speech (TTS) products, according to company’s CTO Marcel Riedi. The male voice brings the total number of languages in the SVOX TTS portfolio to 25 and the number of voices to 36. The new voice is offered across the whole range of TTS solutions based on unit selection technology, namely SVOX Automotive TTS (Standard, Pro, and Plus configurations) and SVOX Mobile TTS (Standard and Pro configurations).

Martin Reber, SVOX CEO, noted, “Arabic is one of the most spoken languages in the world and an official language in 25 countries. As the region grows in prosperity, customers of SVOX are getting more and more interested in the language being accommodated on their automotive and mobile products.”

Apple has a (well-hidden) text-to-speech option in its Preview software (delivered with the Snow Leopard OS)

To create a recording from text in **Apple’s** Preview application to be played in iTunes, once the capability is activated, one can highlight the text, control click it (or right-click it) to reveal the contextual menu, and chose “Add to iTunes as a Spoken Track.” Mac OS 10.6 creates this as an AAC file. To enable this functionality on a Mac computer running OS 10.6 Snow Leopard, in your active application or in the Finder, select SERVICES – SERVICES PREFERENCES. Then, under TEXT, click the checkbox for ADD TO ITUNES AS A SPOKEN PODCAST. Once the preference setting is changed, the above procedure will work.

HP, Real Soft, and USAN using GM Voices for prompt recording

GM Voices provides professionally recorded voice prompts and greetings for telecom applications worldwide. GM Voices is now offering next-day turnaround on small recording orders.

The company announced the deployment of 25 (and counting) international voice personas for HP’s global call center network. Voice personalities created for HP include Neutral Spanish, French Canadian, Mandarin, Cantonese, and many more. HP currently employs more than 23,000 agents at 92 global call center sites. “I personally say we do all languages,” said Senior Account Manager Theresa Lelux. “No matter how obscure or specialized, we can likely connect with a talented voice actor and translator.”

Real Soft, Inc. (RSI), a global provider of IT solutions, recently won new IVR business from **Blue Cross Blue Shield of Delaware** and **Duane Reade Pharmacy** after pitching their platform with GM Voices’ VisionClips. Guy Miller, VP of RSI’s Speech Enterprise Division, said, “GM Voices truly brings our work to life in the eyes of our prospects. Functionality and great sound are hallmarks of the best applications.”

GM Voices also announced a sales partnership with USAN, a provider of call center solutions. “We don’t step into the recording studio until we’ve studied the client’s brand and target market,” said Marcus Graham, CEO of GM Voices. “USAN clients already have well-established brands. We just ensure that their brands are heard clearly in every phone call.”

Frost & Sullivan comments on “behavioral biometrics,” including voice verification

Biometrics is usually considered to identify individuals through aspects of their physiology. However, in a report on Strategic Assessment of Behavioral Biometric Technologies, **Frost & Sullivan** includes aspects of the individual’s behavioral patterns as a biometric indicator. One could argue that some aspects of a voice reflect the resonances of the vocal tract, a physical characteristic, but certainly we all have speaking habits (e.g., accents) that are behavioral.

In its report, Frost & Sullivan does include voice verification, noting that it is related to the unique geometry of the speaker’s vocal tract, including vocal tract length, ratio of larynx to sinuses cadence, pitch, tone, frequency, range, and duration of voice. The report indicates favorably that the technology requires little specialized hardware beyond a microphone, and that it is gaining in popularity.

Nice Systems introduces multi-channel interaction analytics to extract insights from customer touch points beyond voice

Nice Systems has added multi-channel analytics to the speech analytics capabilities of its recording systems (SSN, September 2009, p. 16). The company’s “Interaction Analytics” of its Nice SmartCenter suite now allows organizations to identify and analyze customer behavior patterns across a broad variety of available contact channels, from email and chat to telephone and social media. Nice’s multi-channel Interaction Analytics can automatically correlate a customer email requesting a product upgrade with product feedback communicated via social media, and then ensure that when the same customer calls the contact center, an agent they talk to automatically knows about those recent activities.

Market research firm predicts global automotive telematics shipments to be included in 84.6% of all cars by 2016

Telematics refers to the solutions and applications built on top of information content flowing via wireless communication to and/or from the auto. Beyond accident alerts and remote diagnosis, telematics functions span from the wireless integration of third-party devices, to navigation and Location-Based Services (LBS) updates, to theft detection, to engine control software revisions.

Global shipments of automotive telematics systems are set to rise to 68.4 million units in 2016 (84.6% of all cars shipped in 2016), up from 19.3 million in 2008, according to **iSuppli Corp.** Aftermarket shipments of telematics systems will rise to just under 16 million units in 2016, up from slightly less than 5 million in 2008. For drivers, telematics can enhance safety, convenience, and connectivity. For car OEMs, telematics can add to and improve car functionality and reduce warranty and after-sales costs.

The United States in 2009 accounts for about half of the global installed OEM telematics market. By 2016, the United States will account for about one third of the worldwide market as shipments in other regions grow.

Two major types of telematics solutions are being offered on the market: embedded systems, whose functionality is integrated into the head unit of cars; and mobile-device oriented systems, which use a wireless product like a cell phone to communicate information. Both embedded and mobile device telematics systems are on rapid growth paths. Worldwide OEM embedded telematics systems will grow from nearly 4.8 million units in 2008 to more than 26.8 million systems in 2016; and global OEM mobile device telematics systems will grow from 9.5 million units in 2008 to more than 41.5 million systems in 2016, the research firm forecasts.

3M introduces physician dictation on BlackBerry and Windows Mobile smartphones

3M Health Information Systems has released 3M Mobile Dictation Software, a powerful new application that extends 3M’s dictation, transcription, and speech recognition solutions. (3M has used Philips speech recognition in the past, now supplied by **Nuance**.) Available on the BlackBerry or Windows Mobile platforms, the software offers physicians the freedom of using a single device for phone, email, and dictation, and provides “anytime, anywhere” access via Wi-Fi or 3G wireless service. 3M Mobile Dictation includes security features that fulfill HIPAA and hospital-specific guidelines for encryption and authentication.

3M Mobile Dictation, a module of the 3M Mobile Documentation System, provides always-on connectivity, eliminating the need to synchronize smartphones to a dictation system. The software makes it possible for physicians to view patient lists, search patient IDs, and display the most current patient information on the smartphone screen.

The software integrates with major hospital dictation and transcription solutions. Good sound quality allows integration with speech recognition applications, including 3M SyncStream Intelligent Language Processing Software.

BigHand replaces WinScribe dictation solution at Australian law firm, WinScribe announces award for best legal dictation system

BigHand supplies a digital dictation solution for law firms, using **Nuance** technology (SSN, December 2009, p. 15). BigHand announced that commercial law firm **SR Wallace & Wallace** in Mackay, Queensland, Australia, has rolled out BigHand Digital Dictation to all legal and support staff across the organization. The firm replaced a **WinScribe** Dictation system with the BigHand solution, according to BigHand. Brett Johnson, Managing Partner, SR Wallace & Wallace, cited a number of technical problems and deficient technical support from WinScribe.

Ironically (or perhaps less than a coincidence), WinScribe announced in December that *ALB Legal News* in Australia recently announced the winners of their Service Provider Awards for 2009. Readers voted for WinScribe, market innovator in digital dictation, transcription, and workflow management solutions as Dictation/Transcription Provider for 2009. The award, announced November 19, 2009, was voted on by the readership of *ALB Legal News* for products that end users say make their working lives easier. “WinScribe Dictation v. 4.0, our newest version, has evolved purely from customer feedback,” explains Matthew Weavers, CEO, WinScribe.

Motorola introduces hands-free voice-directed mobility solution for the warehouse

Motorola announced a new rugged mobility solution aimed at improving the productivity of warehouse workers—the WT4090 Voice-only Wearable (VOW) Terminal and the RCH50 Rugged Cabled Headset. This rugged solution can replace paper-based processes with voice-directed solutions in demanding enterprise environments—in the warehouse, in the freezer, on the loading dock or on the factory floor—resulting in real-time processing, reduced errors, and improved decision making. Gerald McNerney, vice president of mobile computing product management, Motorola Enterprise Mobility Solutions, said, “The hands-free, voice-directed WT4090 VOW and RCH50 offer warehouse workers support for voice-directed applications for increased productivity and streamlined operations.”

Bill Anderson, partner, **Genesta**, said, “Purpose-built for industrial voice recognition applications, the RCH50 can withstand tough environments, and our SyVox customers are impressed with its crisp audio and consistency of recognition.”

Grundig Business Systems dictation microphone features support for speech recognition software and low energy consumption

The Digta SonicMic II dictation microphone from **Grundig Business Systems** is the latest microphone in a series of offerings. The handheld microphone has physical controls for dictation systems. The Digta SonicMic II is able, for example, to control **Nuance**'s Dragon NaturallySpeaking speech recognition software via mouse control, slide switch and function keys. The new version's sound quality is also improved, and it's ideally suited for speech recognition purposes, according to the company.

The user has full control of the running PC application via trackpoint mouse control and mouse buttons. The Digta SonicMic II's mouse control now works independently of the regular PC mouse. This feature allows for individual adjustment of SonicMic II's cursor speed, while the PC mouse remains the same. Compared to its predecessor, the Digta SonicMic II is especially distinguished by its low power consumption, contributing the product's durability.

Ginger Software uses text-to-speech as part of software to aid in writing

Ginger Software's Ginger is a tool designed to help the estimated 43.5 million people in the United States with dyslexia. The software is able to decipher the creative spelling errors made by people with dyslexia, allowing them to have confidence in their writing skills. Ginger is an enhanced correction program that corrects multiple spelling and grammatical errors within a sentence at once. Text-to-speech enables users to hear sentences before and after correction, as well as alternative corrections and their definitions. Each word is highlighted as spoken, and personal settings allow students to change the speed, volume, and voice used. Ginger Software uses **Acapela** for text-to-speech technology, according to a spokesperson.

Ginger now works with Internet Explorer. Ginger is also compatible with Microsoft Word and Outlook. “Ginger gives people the freedom to write more efficiently and independently,” said Yael Karov, founder and CEO of Ginger Software. “Writing emails or socializing on Facebook and Twitter can be very intimidating for someone with writing or spelling difficulties. Ginger levels the playing field and allows everyone to participate.”

Aircraft firm chooses Adacel for cockpit voice control system

Adacel is a developer of speech recognition technology; voice-activated control systems; operational Air Traffic Management software; and simulation systems for aircrew, air traffic control, airport vehicle operators, and security training. The company announced a contract award from **Alenia Aermacchi S.p.A.** to integrate Adacel’s Voice Activated Cockpit system into the Aermacchi M-346 aircraft, an advanced fighter trainer. Adacel will supply both the black box hardware and the advanced software required to integrate their speech recognition into the M-346 avionics package. The Adacel Voice Activated Cockpit System uses continuous, speaker-independent recognition.

Adacel’s Direct Voice Input (DVI) control system is designed to reduce pilot workload and cockpit distractions. The advanced speech recognition features developed by Adacel in effect enable the pilot and the aircraft to talk to each other, automating many tasks that would otherwise require multi-step manual input from the pilot.

The Aermacchi M-346 is tailored to train pilots to fly new-generation combat aircraft and helps reduce the flight hours on the more expensive aircraft. The aircraft has avionics modeled upon those of new generation military aircraft such as Eurofighter, Gripen, Rafale, F-22, and the F-35 Lightning II. The M-346 is one of a number of voice-activated aircraft programs the company has supported, including the F-35 Lightning II and Boeing Apache.

Barnes & Noble Nook e-book reader said to have hooks for future speech recognition

Apparently, the **Barnes & Noble** Nook e-book reader had been hacked to grant “explorers” root access and learn more about its internals. Hacking the Nook apparently only requires knowledge, a screwdriver, a microSD card reader, fingernails, and a computer running Linux. The process requires changing a boot file for the OS installed on an internal microSD card, then using the Android development kit supplied by **Google** to tweak a few settings. According to reports, those who have looked at the list of files on the Nook say that it shipped with support for a speech recognition engine.

Speaker identification plays role in Jakarta trial

An article in the *Jakarta Post* said that an expert witness testifying in the high-profile murder trial of former anti-graft leader Antasari Azhar claimed that a recording of an incriminating conversation between Antasari and his friend Sigid Haryo Wibisono, a businessman, was authentic. The expert, M. Nuh, an IT analyst from the National Police’s forensic laboratory, who was asked by police investigators to analyze and transcribe the recording, said he was sure the voices in the recording were Antasari’s and Sigid’s after comparing voice samples. “Our voice recognition analysis found that 20 words said by Antasari in the recording were identical to his voice sample,” he said, adding that according to FBI standards, a 20-word match was sufficient to positively identify someone’s voice in a recording. In parts of the recorded conversation, Antasari and Sigid rehearsed scenarios to kill businessman Nasruddin Zulkarnaen, prosecutors said. The defense, of course, challenged the testimony, citing, among other things, missing words in a transcript.

Google “translated search” allows finding information from sites written in other languages

In a Web posting by Maureen Heymans, Technical Lead, and Jeff Chin, Product Manager, **Google** introduced a new “translated search” tool in the Search Options panel that makes finding and reading content written in other languages easier. Translated search helps one find information from sites written in other languages. The feature has been available in Google Translate, but is being integrated fully into Google search. To search the web in another language, click “Show Options” at the top of the search results page and select “Translated search.” Google uses software to select the best language(s) to translate a search query into and then returns translated results from those pages.

Lingo Media subsidiary Parlo to train China Southern Airlines' cabin attendants in spoken English using speech recognition

Lingo Media Corporation announced that its wholly-owned subsidiary, **Parlo Corporation**, a training and assessment platform and its joint venture partner, **Lion King Education (International) Group (LKE)**, have signed **China Southern Airlines** as the first corporate training client for its English language training platform. China Southern is the largest international airline operating in China, and the world's fifth largest airline by passengers carried.

The service will be used to train more than 5,000 cabin attendants and over 1,200 ground service staff in job-specific spoken English. The Parlo solution will repurpose China Southern's training manuals to create nearly 150 lessons that teach cabin crews to deliver English language announcements and to handle conversations with passengers and international support staff. The solution integrates Parlo's proprietary speech recognition system to train and test pronunciation. Animated avatars take the place of live instructors.

Toshiba develops a speech-to-speech translation system that can run entirely on a mobile phone

Toshiba has developed a trilingual translation system with speech recognition and synthesis that can be installed in a cell phone to translate between Japanese, English, and Chinese. The embedded software doesn't offload processing to a server on the network, but performs the task inside the phone. That makes it quicker and avoids potentially costly data roaming charges when used overseas. The software is a reduced version of a PC application already sold by Toshiba in Japan.

The software first employs speech recognition to determine both the language and what has been said. It then uses one of two translation methods—statistical machine translation or rule-based machine translation—to parse the sentence and break it down into its components. The translation is then performed and a text-to-speech engine speaks the result.

Statistical machine translation works on probability and word order to determine the construction of a sentence while the rule-based method utilizes a database of hundreds of sentences and phrases to extract the meaning. The PC software uses both methods simultaneously but a cell phone lacks enough processing power to run both so statistical machine translation is used for English to Chinese and Chinese to English and Japanese while rule-based machine translation is used for the others.

Speech recognition techniques to check your heart?

An article in the UK newspaper *The Guardian* discusses the use of technology derives from speech recognition technology that may be that a better predictor of heart problems than standard tests. The computer program analyzes the patient's electrocardiogram looking for subtle features. A team of researchers at **MIT** and the **University of Michigan** analyzed a large data-set of 24-hour electrocardiogram recordings collected at a Boston hospital. The researchers developed a way to analyze how the shape of the electrical waveform varies, a measure they dubbed morphological variability. At the heart of the approach are mathematical techniques used in speech recognition that allow researchers to compare variability in individual beats. If there is significant variability between individual beats, that signals a problem.

The team then applied their algorithm to a second set of electrocardiogram recordings and found that patients with the highest morphological variability were six to eight times more likely to die from a heart attack than those with low variability. They concluded that it consistently predicted as well or better than the indicators commonly used by physicians.

TD Canada Trust and TD Waterhouse add web accessibility features

TD Canada Trust and **TD Waterhouse** launched a new tool to help visually impaired customers with their online banking. EasyWeb and WebBroker feature a new web accessibility tool in the form of a toolbar that allows users to adjust font sizes, change color contrast, magnify all or a portion of the screen, and have text read aloud with adjustable speed and volume control. Developed by **IBM Research** in 2003, the toolbar significantly enhances the usability of online banking services for customers with low vision.

The IBM Human Ability and Accessibility Center focuses on technology solutions to address the issues associated with disability, aging, and low literacy. Part of IBM Research, the Center is a worldwide organization that works with governments, collaborates with partners, and delivers solutions to clients in the Americas, Asia, and Europe.

Tigal offers voice control module for robots

Robots for hobbyists can be quite complex. The ROBONOVA-I from **HiTec** can walk, run, do flips, cartwheels, and dance moves, and is programmable. The robot is 12.5" high and sells for about \$1,300.

TIGAL has launched a voice control unit for ROBONOVA-I and ROBOZAK robots. Tigal's VRbot module provides users with 26 built-in speaker-independent commands that allow basic control of robot movement without any programming. Users can customize up to 32 user-defined speaker-dependent commands and assign these to specific movement sequences of the robot using Windows-based GUI software included with the module. The module has been designed to fit inside the chest cover of the robots.

New draft of VoiceXML 3.0 spec published

The fourth Working Draft of VoiceXML 3.0 has been published at <http://www.w3.org/TR/2009/WD-voicexml30-20091203/>. The main difference from the previous draft is addition of run time control and Speaker Identification and Verification modules.

New working draft of W3C "Multimodal Architecture and Interfaces" available

The sixth Working Draft of "Multimodal Architecture and Interfaces" has been published at www.w3.org/TR/2009/WD-mmi-arch-20091201. Kazuyuki Ashimura, W3C Multimodal & Voice Activity Lead, reports that the document as a whole has changed significantly and the group welcomes review. The main changes from the previous draft are (1) clarifying the relationship to EMMA, (2) simplifying the architecture constituents, (3) adding a description on HTTP transport of lifecycle events and (4) adding an example of handwriting recognition modality component.

Call Genie to design, deploy and maintain an end-user facing directory assistance website for TDC Denmark

Call Genie, which recently completed a financing (p. 36), announced that it has entered into a three-year agreement to design, deploy and maintain an end-user facing directory assistance website for **TDC Denmark**, the leading telecommunications company in Denmark. TDC has contracted with Call Genie to develop and implement a comprehensive new database to supplement TDC's current 118 database (comparable to "411" in North America). This deal doesn't include a telephone component, the arena that most associate with Call Genie.

Best wishes from Speech Strategy News for the New Year!

May all go exceptionally well for you and yours.

- Bill

Financial Notes

SpinVox takeover by Nuance rumored

The UK's *Sunday Times* reported on December 13 that **SpinVox**, the voicemail-to-text company (p. 13), was expected to accept a GBP92 million (US\$150 million) takeover offer from **Nuance Communications**. The article also indicated that SpinVox has been granted an extension until next year to pay back a loan of GBP30 million (US\$49 million) while negotiations are taking place. Nasdaq-listed Nuance has a market value of \$4.2 billion.

SpinVox previously received £100 million of backers' money, but the use of agents to transcribe part of some messages has apparently been expensive. In August, Tisbury, a fund manager, extended a £30m bridge loan. The loan has been rolled over to January to allow the takeover to be completed. SpinVox is projected to earn £7 million next year and £30m in 2011 after signing contracts with several mobile operators. Unaudited accounts for 2008 report a pre-tax loss of £49m.

Nuance officers' bonuses reported

Nuance Communications filed a report on bonuses for some executives on December 14, 2009, with amounts that perhaps reflect the more conservative remuneration policies of technology firms relative to

financial firms. The Compensation Committee of the Board of Directors of Nuance approved compensation awards to be granted pursuant to the Company Bonus Program for fiscal 2009 to the following executive officers of the Company in the amounts set forth below (part in cash and part in restricted stock options that require continued employment):

Name and Title	Total Bonus
Paul A. Ricci Chief Executive Officer	\$ 550,000
Thomas L. Beaudoin Executive Vice President and Chief Financial Officer	\$ 210,000
Steven G. Chambers Executive Vice President and Worldwide Sales and Chief Marketing Officer	\$ 250,000
Jeanne F. McCann Executive Vice President and General Manager, Healthcare Division	\$ 125,000

Call Genie announces closing of \$4.0 million convertible debenture financing

On December 29, **Call Genie Inc.** (TSX: GNE), a global provider of local mobile search and advertising solutions (p. 35), announced that it has distributed an aggregate amount of \$4.0 million principal amount of convertible secured debentures including the \$2.5 million aggregate amount of the first tranche previously announced on November 12, 2009. Under the terms of the debentures, outstanding principal will bear interest at a rate of 12% per annum, payable semi-annually, and the debentures will mature on October 31, 2012. At the option of the Debenture Holder, principal and accrued interest under the debentures may be converted into common shares of the Company at a conversion price of \$0.50 per share. The Company has granted a security interest to Debenture Holders under a general security agreement covering all of the assets of the Company. In addition, each Debenture Holder received 480 warrants per \$1,000 principal amount of debentures. Each warrant entitles the holder to purchase one common share of the Company, at an exercise price of \$0.225, at any time prior to October 31, 2011. The debentures have not been, and will not be, registered under the *United States Securities Act of 1933*.

Avaya completes acquisition of Nortel enterprise solutions

On December 18, **Avaya** announced the successful completion of its acquisition of **Nortel Enterprise Solutions (NES)**. The acquisition gives the company added scale, resources, and expertise to deliver a portfolio of solutions and services around the globe. "The completion of this acquisition represents another major step in Avaya's evolution and growth in the communications industry," said Kevin J. Kennedy, president and CEO, Avaya. "Avaya and Nortel Enterprise Solutions share a common vision for the future of business communications. By combining our complementary technology portfolios, deep industry specific domain expertise, sales channels and customer bases, the new Avaya will redefine business communications and help customers to reduce costs, simplify operations and increase their business agility."

Approximately 6,000 NES employees have joined Avaya, including 25 top managers. Joel Hackney, previously president, Nortel Enterprise Solutions, joins the Avaya Executive Committee as senior vice president and president, Avaya Government Solutions and Data.

For Further Information on Products Mentioned in this Issue

Company	Location	Product Mentioned	Contact info
3M Health Information Systems	Salt Lake City, UT	Healthcare software and services	(800)367-2447; www.3Mhis.com
Abvio	--	Fitness products	www.abvio.com

Companies mentioned in this issue

Company	Location	Product Mentioned	Contact info
Aculab	Milton Keynes, UK	Telephony boards with speech recognition	+44 1908 273 800; www.aculab.com
Adacel Inc.	Orlando, FL	Defense and air traffic control solutions	(407)581 1560; www.adacel.com
Alpine Electronics, Inc.	Torrance, CA	Automotive electronics	310-326-8000; www.alpine.com
Apple	Cupertino, CA	Personal computers, music players, wireless phones	www.apple.com
Aspect Software	Chelmsford, MA	Telephone self-service system with speech recognition	(978)250-7900; www.aspect.com
AT&T	San Antonio, TX	Telecommunications services	www.att.com; www.wireless.att.com
Avaya Inc.	Basking Ridge, NJ	Enterprise telephony solutions	(908)953-6000; www.avaya.com
Barnes & Noble	Lyndhurst, NJ	e-book reader	www.barnesandnoble.com
Best Buy	Minneapolis, MN	Retailer	www.bestbuy.com
BigHand	Chicago, IL (London, Sydney, Toronto)	Professional transcription options	(312)893-5906; www.bighand.com
Call Genie	Calgary, AL, Canada	Speech recognition solutions for the directory services industry	(403) 268-0411; www.callgenie.com
Cisco	San Jose, CA	Internet infrastructure and IP telephony	(800) 553-6387; www.cisco.com
Confermit	Oslo, Norway	Hosted customer feedback surveys	+47 21 502 500; www.confermit.com
Dialogic	Montreal, Canada	Communication products	(514)745-5500; www.dialogic.com
Digium	Huntsville, AL	Telephone interface cards and support for open-source Asterisk PBX	(256)428-6000; www.digium.com
Ditech Networks	Mountain View, CA	Voice quality and voice access solutions	(650)623-1300; www.ditechnetworks.com
DMG Consulting	West Orange, NJ	Contact center consulting	(973)325-2954; www.dmgconsult.com
DynaVox Mayer-Johnson	Pittsburgh, PA	Assistive communication devices	(412)381-4883; www.dynavoxtech.com
Ford	Detroit, MI	Vehicles and informatics	www.ford.com
Frost & Sullivan	Mountain View, CA	Voice messaging report	(415)961-9000; (415)961-5042(fax); www.frost.com
Genesta	Rockwall, TX	Industrial and logistics solutions	(972)771-1653; www.genesta.com
Genesys Telecommunications Laboratories (Alcatel-Lucent subs.)	Daly City, CA	Call routing and contact center solutions	(888)GENESYS; www.genesyslab.com
Ginger Software	—	Spelling error correction software	www.gingersoftware.com
GM Voices	Alpharetta, GA	Recorded prompts and audio	(770)752-4500; www.gmvoices.com
Google	Mountain View, CA	Voice and directory search	(650)253-0000; www.google.com; www.google.com/mobile; www.grandcentral.com
Grandsys	China	Call recording	www.grandsys.com/en
Grundig Business Systems	Germany	Dictation solutions	www.grundig-gbs.com
Hitec Robotics	Poway, CA	Hobby robots	(858)748-6948; www.robonova.com
IBM Research	Yorktown Heights, NY	Speech recognition research	(914)945-3000; www.ibm.com
ICR	Saltaire, UK	IVR solutions	+44 1274 821111; www.icr3s.co.uk
Intel Corporation	Santa Clara, CA	Computing and communications chips, solutions, and platforms	www.intel.com
Interact Incorporated	Lincoln, NE	Communications solutions	(402)476-8786; www.iivip.com
Interlancompu	Quito, Ecuador	IT consulting	+593 2 2526585TELEFONO; www.interlancompu.com

Companies mentioned in this issue

Company	Location	Product Mentioned	Contact info
Intermec Technologies Corp.	Everett, WA	Supply chain management and data collection devices	(425) 348-2600; www.intermec.com
INX Inc.	Canton, MA	Unified communications and data center virtualization solutions	(339)502-5300; www.inxi.com
iSuppli Corp.	El Segundo, CA	Electronics market research	(310)524-4007; www.isuppli.com
J.D. Power and Associates	Westlake Village, CA	Surveys and studies	www.jdpower.com
Jingle Networks, Inc.	Bedford, MA	Free Ad-supported Directory Assistance	(877)754-6453; www.free411.com; www.jingleconnect.net
KLAS Enterprises	Orem, UT	Healthcare vendor evaluation	(800)920-4109; www.KLASresearch.com
Lingo Media Corporation	Toronto, ON, Canada	English language learning	(416)927-7000 ; www.lingomedia.com
Loquendo	Turin, Italy	Speech technology licensing	+39 011 291 3111; www.loquendo.com; Developers: www.loquendocafe.com
LumenVox LLC	San Diego, CA	Speech recognition engine and development tools	(858)707-0707; www.lumenvox.com
Luna Project	European Union	Research consortium on Spoken Language UNderstanding in multilinguAI Communication Systems	www.ist-luna.eu
MedQuist Inc.	Mount Laurel, NJ	Medical document creation	(800)233-3030; www.medquist.com
Microsoft Corporation	Redmond, WA	Various applications and products using speech technology	(206)454-2030; www.microsoft.com/speech
Morgan Stanley	New York, NY	Investment firm	www.morganstanley.com
Motorola	Downers Grove, IL	Mobile phones, portable devices	(630)353-8000; www.motorola.com
Nexidia	Atlanta, GA	Audio content search	(404)495-7220; www.nexidia.com
Nice Systems	Ra'anana, Israel	Multimedia analytics	+972 9 775-3777; www.nice.com
Nortel	Bohemia, NY	Telephony and networking systems for service providers and enterprises	1-800-4NORTEL; www.nortelnetworks.com/solutions
Novauris Technologies Ltd	Cheltenham, England	Speech recognition technology	+44 1242 678581 (UK); (530)753-1160 (US); www.novauris.com
Nu Echo Inc.	Montréal, Canada	Speech application consulting	(514)861-3246; www.nuecho.com
Nuance Communications	Burlington, MA	Speech technology, applications, and services	(617)428-4444; www.nuance.com
O'Reilly Automotive, Inc.	Springfield, MO	Automotive aftermarket parts	(417)862-2674; www.oreillyauto.com
Optus Mobile	Australia	Mobile telephone service	www.optus.com.au
Paisley Group	Castle Rock, CO	Directory assistance consulting	(303)688-1666; www.thepaisleygroup.com
ParkMobile	Diemen, The Netherlands	Electronic parking	+31 20-560 10 50; www.parkmobile.com
Parlo	Toronto, ON, Canada	Training and assessment	(416)927-7000; www.parlo.com
Persay	Woodbury, NY, and Tel Aviv, Israel	Speaker authentication technology	1(516)677-7291; +972-3-7678666; www.persay.com
Plum Voice	New York, NY, and London, UK	Customized IVR solutions	1-800-995-PLUM; www.plumvoice.com
Primordial Design, Inc.	—	Communication, workflow, quality management and analytics applications for radiology	(800)398-1140; http://primordialdesign.com
Promptu	Menlo Park, CA	Voice-activated search services for cable and mobile	(650)859-5800; www.promptu.com
ReceptionHQ	Phoenix, AZ	Answering services provider	(866)883-3499; www.receptionhq.com
Remego Consulting	Singapore	IVR solutions	+65 6754 7366; www.remego.com

Companies mentioned in this issue

Company	Location	Product Mentioned	Contact info
RoboTech srl	Pisa, Italy	Entertainment and educational robotics	+39 0587 672027; www.robotechsrl.com
Sabse Technologies	Mountain View, CA	Telephony-as-a-Service (TAAS) platform	www.sabsebolo.com
Sandata Technologies	Port Washington, NY	IT solutions for home healthcare and social services	(516)484-4400; www.sandata.com
Sensus	Hillerød, Denmark	Accessibility research	www.sensus.dk
SpeechCycle	New York, NY	Automation of telephone customer care	(646)792-2720; www.speechcycle.com
SpinVox	Marlow, UK	Voicemail-to-text service	+44 020 7965 2000; www.spinvox.com
Strategy Analytics	Newton, MA	Market reports	617 614-0700; www.strategyanalytics.net
SVOX AG	Zurich, Switzerland	Speech recognition and text-to-speech technology	+41 43 544 06 23; www.svox.com
Syntellect Inc. (subs. of Enghouse)	Phoenix, AZ	Voice processing platforms	(602)789-2800; www.syntellect.com; www.voiyager.com
TantaComm	Middleton, WI	Call recording solutions	(800)444-8522; www.tantacomm.com
TDC Denmark	Denmark	Telephone service provider	www.TDC.com
TelcoBridges	Boucherville, QU, Canada	Telecom hardware and software	(450)655-8993 ; www.telcobridges.com
Telephonetics VIP (unit of Telephonetics plc)	Hemel Hempstead, UK	Contact center solutions	www.telephoneticsvip.co.uk
Tellme (Microsoft subs.)	Mountain View, CA	Voice application hosting and services	(650)930-9000; www.tellme.com
TelStrat	Plano, TX	Contact center solutions	(972)543-3500; www.telstrat.com
The Joint Commission	Oakbrook Terrace, IL	Healthcare certification	(630)792-5000; www.jointcommission.org
Tigal	Austria	Distributor of technical products	+43 2231/68 347; www.tigal.com, www.veear.eu
Toshiba	Japan	Telephony solutions	www.solutions.toshiba.com
USAN	Norcross, GA	Call center hosting	888-676-1112; www.usaninc.com
Utopy	San Francisco, CA	Call center speech mining	(415)621-5700; www.utopy.com
Verbio	Barcelona, Spain	Speech technologies	+34 93 444 79 79; www.verbio.com
Verizon Business	USA	Managed and/or hosted enterprise solutions	1-877-297-7816; www.verizonbusiness.com
Vlingo, Inc.	Cambridge, MA	Voice-powered interface for mobile phones	(617)871-2987; www.vlingomobile.com
Vocollect, Inc.	Pittsburgh, PA	Portable industrial speech recognition system	(412)829-8145; www.vocollect.com
Voitra Technologies	—	Live and automated agent software for VoIP networks	(877)486-4872; www.voitra.com
Volt Information Sciences, Inc.	New York, NY	Workforce, information, and telecommunications solutions	www.volt.com
Voxeo	Orlando, FL	Voice hosting solutions	(407)418-1800; www.voxeo.com
Voxware, Inc.	Princeton, NJ	Industrial speech recognition	(609)514-4100; www.voxware.com
W3C Multimodal	—	Standards effort	www.w3.org/TR/emma/
W3C Voice Browser Working Group	—	Standards effort including VoiceXML	dsr@w3.org; www.w3.org/voice
West Corporation	Omaha, NE	Outsourced communication solutions	www.west.com
West Interactive (unit of West Corp.)	Omaha, NE	Out-sourcing of customer contact solutions	(402)963-1300; www.westinteractive.com
WinScribe	Chicago, IL	Dictation solutions	(866)494-6727; www.winscribe.com
Zetes	Brussels, Belgium	Voice-enabled warehouse solution	+32 2 728.37.11; www.zetes.com

Google signs on as principal sponsor of the Mobile Voice Conference 2010

Introducing the decade of the Mobile Internet with a critical role for speech technology

The **Applied Voice Input/Output Society (AVIOS)** and Bill Meisel’s **TMA Associates** are organizers of the third annual *Mobile Voice Search Conference* (previously the Voice Search Conference), to be held April 22-23, 2010, at the Hyatt Fisherman’s Wharf in San Francisco, California. **Google** is a Principal Sponsor, and **Ditech, Loquendo, Novauris, and West** are supporting sponsors. A major investment firm named this decade the decade of the Mobile Internet, and usability will determine the degree to which the PC version of the Web translates to mobile phones (see Editors Notes, p. 4).

The conference examines the growth in the use of speech technology (speech recognition in particular) driven by wide adoption of mobile phones. Mobile phones are motivating a different style of application, helping address the limitations of a Graphical User Interface and text entry on the small device. The growing familiarity and credibility of mobile speech applications is also creating new expectations in other environments, such as call centers.

The Mobile Voice Conference is the first to fully examine the impact of this paradigm shift. For more information, see www.mobilevoiceconference.com.

I wish to **subscribe** to *Speech Strategy News* for one year (12 issues), payable in US\$ on US bank—

Destination	Individual rates			Corporate rates*		
	Paper+Web+Archives**	Web+Archives**	Paper only	Paper+Web+Archives**	Web+Archives**	Paper only (5 copies)
U.S. & Canada	<input type="checkbox"/> \$495	<input type="checkbox"/> \$425	<input type="checkbox"/> \$425	<input type="checkbox"/> \$1,895*	<input type="checkbox"/> \$1,495*	<input type="checkbox"/> \$1,495
International	<input type="checkbox"/> \$525	<input type="checkbox"/> \$425	<input type="checkbox"/> \$480	<input type="checkbox"/> \$1,895*	<input type="checkbox"/> \$1,495*	<input type="checkbox"/> \$1,495

* *Corporate subscriptions*: Unlimited users within a corporation for Web+Archives; one paper copy mailed in addition if Paper+Web+Archives elected. Five paper copies, sent separately or as a group, if Paper Only elected.

** *Web + Archives*: Searchable current and back issues. Issue on Web can be printed.

I wish to receive **selected news alerts** (less than 2 per month) by email at no additional cost. Email address for alerts (available with all subscription categories): _____

Please send information on your **consulting**.

Name: _____

Company: _____

Address: _____

City, State _____

ZIP/Postal code _____

Country _____

Email (required for email alerts or a Web subscription): _____

Phone: _____

Check enclosed, payable to TMA Associates (in U.S. \$ on a U.S. bank).

Invoice me.

Charge my—

Visa MasterCard American Express

Card # _____

Expiration date: _____

Signature: _____

SSN199

Copyright TMA Associates 2009; All rights reserved. TMA Associates, P.O. Box 570308, Tarzana, CA 91357-0308 USA. Tel: (818) 708-0962. Fax: (818) 232-0368, or go to www.tmaa.com/subscribetossn.

Speech Strategy News is published twelve times per year by TMA Associates, Editor: William S. Meisel. Trademarks mentioned in this publication are the property of the companies mentioned; they are used editorially. The material herein is based on data from sources believed to be reliable, but is not guaranteed as to accuracy and does not purport to be complete. From time to time, the author or TMA Associates may have consulting assignments, advisory positions, own stock, or have other business relations with organizations in speech recognition and associated areas, including companies discussed in this newsletter. *Speech Strategy News* is a trademark of TMA Associates.