



## WHITEPAPER

# Hosted and Premise-based Dictation and Transcription Systems using SPART<sup>®</sup>

by:  
Veeru Ramaswamy, PhD  
Chief Technology Officer  
Vianix, L.L.C.  
September 2, 2009

## **Hosted and Premise-based Dictation and Transcription Systems using SPART<sup>®</sup>**

### *Abstract*

Dictation and Transcription (DT) workflow systems for the Healthcare, Legal and Insurance industries comprise a multi-billion dollar market. There are two ways DT systems are implemented and deployed: Premise-based and Hosted/ASP/On-demand. Historically, deployments were exclusively Premise-based systems, where the servers are deployed in the organization's existing IT. Today, in the Hosted/ASP/On-demand case, the servers are part of data-centers or IT-cloud hosted by an ASP or network service provider.

The Premise-based system usually involves buying hardware (off-the-shelf or specialized systems) along with different software components to enable efficient voice and data capture. The captured voice can be aggregated or routed to a local voice server which is then forwarded to the appropriate transcription and/or database servers inside or outside the local network.

In a hosted environment, also known as ASP (Application Service Provider) or SAAS (Software-As-A-Service) environment, the data-center or the service provider already has a large IT network in place, servicing different kinds of telephony and IT applications. Only the appropriate software may have to be installed on their servers to aggregate the recording or dictation through Voice over IP networks, typically through web-based applications.

In some cases, organizations also use a mixture of both approaches depending on the cost, scale, security, time-to-market, network speeds/bandwidth and service turn-around times.

In this paper we demonstrate how Vianix's SPART<sup>®</sup> (Speech Processing for Automatic Recognition Technology) technology creates overall bottom-line savings in both approaches.

### **Advantages and Disadvantages of Hosted and Premise-based Dictation and Transcription workflows**

There are advantages and disadvantages to both kinds of systems.

#### *Cost:*

In the premise-based infrastructure, all hardware and software has to be completely managed by corporate IT and telephony divisions. This leads to large initial CAPEX, and to an even greater extent, OPEX as well that can affect and delay the deployment of the dictation and transcription application. In the Hosted case, the service providers typically charge a one-time fee or a nominal monthly subscription fee for the servers and the

transcription service provider can upload or deploy the software application on the leased servers. One has to consider the Total Cost of Ownership (TCO) for both deployment scenarios.

*Scale:*

Depending on how large the customer base or the customer usage is, a Transcription Service Operator (TSO) would prefer to host their service in their own data centers as an ASP/SAAS provider or use their customer's existing corporate infrastructure (Premise-based). For example, in the healthcare market, small clinics or physician's offices do not want the task of maintaining the IT infrastructure on their own and would prefer to outsource this operation to the TSO. In cases where the customers are large hospitals with significant and well maintained IT infrastructure in place, the application is Premise-based within the existing IT (and in-house transcriptionists are used as well). In a mixture of the two deployment models, especially for mid-tier hospitals or a group office, the corporate network use the TSO's infrastructure to transfer voice or data information as a last hop.

*Security:*

Security is a major requirement for the Hosted and Premise-based TSO's. In the healthcare industry, these networks need to follow the HIPAA compliance for preserving the privacy of patient information. So any audio or data with the patient's health and other secure information in it has to be guarded against any malware/intrusion attacks on the system. In the legal and insurance industries, the legal proceedings of any particular user need to be protected against any other party's intrusion into the system. In a Premise-based system it is easier to install specialized software to augment pre-existing security in IT networks, which can then be utilized for voice networks as well. Also, the IT administrators are usually aware of existing attacks into the system. In case of a data-center hosted system using a TSO, the primary concern for any hospital is maintaining HIPAA compliance because voice is being moved outside the internal IT infrastructure.

*Time-to-use:*

For any TSO and healthcare organization, time-to-use is an important factor for their dictation/transcription application to be available in an expedited manner. From a time to use perspective, the hosted model is much faster to deploy than the premise-based model due to SAAS type deployments where the existing servers get uploaded/upgraded with the appropriate software and the existing hardware and network infrastructure is utilized.

*Network speeds/bandwidth:*

Network speeds and bandwidth play a vital role in efficiently transcribing the voice data to text. In Premise-based model with internal transcriptionists, with LAN connectivity being 100/1000-BaseT networks, usually there are no bottlenecks. If the corporate LAN happens to be a narrowband network with large number of dictations queued up on the server, then the efficiency in terms of turn-around times can be affected due to job queues for audio backed-up. In hosted model, where the backend systems are usually fiber-optic networks, it may not be an issue but the connection into the hosted data centers or out of the data

centers to the offsite transcriptionist could be a significant issue. So connectivity plays a very important role in both cases for efficient job turn-around times.

Healthcare professionals, such as physicians, dictate on mobile devices, and sync up with the network (through PC-connected to LAN or WLAN) or perform all their dictation, in both cases, at the end of the day. Networks need to be able to accommodate burst/peak times such as end of day and low-usage times accordingly. Similarly, off-shore transcriptionists may login into a local system during off-working hours in the US. The network bandwidth at both ends needs to provide efficient speed to transfer voice and data.

In both Hosted and the Premise-based models, when the user (physician, lawyer or insurance agent) wants to edit the text on their own, the bandwidth needs to support self-edit mode on the desktops or other devices through which they access the audio and text.

#### *Service Turn-Around Times (TAT):*

This is a very important factor in both hosted and premise-based systems. Physicians, lawyers and insurance agents require their reports be generated quickly. Apart from faster access by transcriptionists, the audio quality is very important for accurately understanding complex professional vocabularies. Also Automatic Speech Recognition engines need to recognize words accurately. In the case of Premise-based system the TAT is shorter, because the transcriptionists are located in the same corporate office. In the case of the Hosted-model where dictation is being transcribed in data centers, the transcriptionists are remotely located and TATs can extend from 24 to 72 hours. Differences in time zones can actually help when the transcriptionists are remotely located, especially when dealing with a large number of queued dictations.

#### *SPART<sup>®</sup> Role in Hosted and Premise-based Solutions*

Users (Doctors, Lawyers, and Insurance agents) dictate into any of the available capture devices (such as handheld voice recorders, smart-phones, PC-microphone, VoIP, or standard telephone). The captured digital voice data is processed through SPART<sup>®</sup> and sent to an ASR engine (SPART<sup>®</sup> is "ASR-agnostic"), with voice and text data then stored in transcription servers, and forwarded either to QC/QA Transcriptionists for editing the reports, or to the users for self-edit. In such as scenario we have shown that SPART<sup>®</sup> has dramatically increased productivity for Transcriptionists and Transcription Service Organizations (TSO's) due to:

- **Cost:** Highly cost effective to transport SPART<sup>®</sup> processed audio since it reduces bandwidth by at least 10-20 times. This is particularly important when Transcriptionists are located in remote locations (usually used with a Hosted model), and more so, outside the country such as Philippines and India.
- **Security:** SPART<sup>®</sup> is inherently secure and so when audio with patients private information is being transported across Public networks, audio cannot be decoded by intruders

- *Time-to-use:* SPART® technology is available in widely used Operating Systems such as Windows including Windows Mobile and Linux along with a broad range of High Level APIs supported on different frameworks such as .NET, JAVA and native C/C++ for various platforms. SPART is also available for specialized DSPs such as ARM based Marvel PXA processors, ADI BlackFin, TI 54x, 55x and 64x infrastructure. This enables SPART® technology to be integrated easily with existing infrastructure or products without having to perform complex programming interfaces.
- *Scalability:* SPART® is multithreaded and supported on most widely used web infrastructures such as .NET and JAVA architectures which are inherently scalable and also using native C/C++ frameworks for mobile platforms.
- *Network Bandwidths:* Due to the fact that audio is being compressed, TSO's can use fewer T1s or OCs to connect to the remote transcriptionists and also to connect several remote locations of physicians, lawyers or insurance agents. The users can transfer the audio to the transcription servers using 2.5G or 3G data networks or any WLAN networks if they are premise-based.
- *Turn-Around Time:* SPART® enables faster TAT due to
  - Faster uploads and downloads by the Transcriptionists for the TSO's.
  - Extreme high voice quality reducing the need for rewinds and fast-forwards for the Transcriptionists to replay audio, and hence increasing their productivity
  - Increased ASR accuracy thereby reducing the number of words to be edited and increasing overall transcriptionist's efficiency.

## Summary

Healthcare, Legal and Insurance Dictation and Transcription is estimated at over \$20 billion dollars in annual revenue, with global reach and rapid Return on Investment (ROI). The TSO's usually deploy one of two models (hosted or premise-based) or a mixture of two models and there are several tens of corporations deploying dictation/transcription workflow applications. As a result, a large volume of digital voice data are captured and transferred around the world.

SPART® plays a significant and crucial role for dictation/transcription workflows for both hosted and premise-based models by improving the reconstructed voice quality and accuracy when used for transcription using ASR engines and thereby improving transcriptionist's productivity and TSO's efficiency.

Vianix' SPART® technology offers the greatest level of voice compression with the highest quality of voice reproduction of any product to date. It provides reduction in the cost of both bandwidth and storage requirements as well as insuring the highest level of voice accuracy. In addition, SPART® is fully compatible with most communication protocols and is, or can be, optimized and customized for integration with any voice

activated system. Because of these qualities, SPART<sup>®</sup> is the one voice compression technology that is ideally suited for storing and transmitting cell phone voice recordings.

For companies involved in developing services using Voice Recording, Vianix offers a complete SPART<sup>®</sup> partner program, which includes:

- Windows and Windows CE ACM codecs, a Software Development Kit for the SPART<sup>®</sup> codec and an application suite, which includes API level access to library classes, functions, and procedures to allow a seamless integration with partner products.
- Means to convert from the SPART<sup>®</sup> proprietary format to a generic format, and vice versa. A customized licensing plan to seamlessly match a partner's business model.

Vianix<sup>®</sup>, SPART<sup>®</sup> and MASC<sup>®</sup> are registered trademarks of Vianix, LLC. All other names and trademarks are the property of their respective owners. For more information on Vianix<sup>®</sup>, visit our website at [www.vianix.com](http://www.vianix.com).