white paper

Choosing the Right Communication Adapter for Speech-Enabled Systems

Executive Summary 1

Overview of the Speech Application Market 2

The Communication Adapter is Key 3

What To Look For 3

Benefits of Eicon's Diva Server V-Series in Speech Applications 5

Going Beyond the Basics 6

Eicon's Diva Server V-Series – Optimized for Speech . . . 7

Conclusion 8

Executive Summary

Advanced speech-enabled applications and systems are changing the way in which we interact with all types of businesses. Moving beyond a few showcase situations, speech technologies are being used in many new and exciting applications, such as voice portals, call center self-service, directory assistance, and voice-activated dialing. With speech applications customers can obtain information or direct transactions using the most intuitive interface - speech. When used in conjunction with IVR and self service applications, speech recognition technology greatly reduces the cost of staffing a call center operation, while reducing call durations and call handling time and increasing call completion rates. Automating some functions through speech recognition technology can significantly reduce costs – even up to 50%, according to one Gartner Group estimate. Speech technologies can provide a return on investment of up six-to-twelve months, making it a must-have on every developer's checklist.

We're experiencing a new wave of development tools and platforms making it easier for customers, system integrators, VARs and other developers to quickly and easily develop new speech applications. As speech-enabled applications continue to grow and evolve, so too do the number of Independent Software Vendors (ISVs) developing these applications. For speech ISVs, choosing the right hardware platform is essential, and ultimately results in reduced costs, faster time to market, and more satisfied customers. Eicon Networks' Diva Server V-Series communication adapters enable developers to create speech applications and systems that meet the current and evolving needs of their customers. The Diva Server V-Series provides the essential elements needed for a speech application solution – while providing the performance and functionality needed at a cost-effective price.

This White Paper discusses what ISVs and system integrators should look for in a communication adapter for speech-enabled systems, and why they must carefully select the right one for their needs.

Overview of the Speech Application Market

In advanced speech applications (sometimes known as Speech-enhanced IVR), automatic speech recognition (ASR) and text-to-speech (TTS) are used together to improve the customer experience over conventional IVR (i.e. voice prompts and DTMF input). Of the two primary speech technologies, speech recognition has seen the greatest acceptance over the past few years. Speech recognition has become mainstream and is being used in applications ranging from airline reservations to stock quotes to creating an appointment on your Microsoft Outlook calendar. The rise in speech recognition applications is due to a great many factors, the most significant being dramatically improved recognition accuracy, increased scalability, and more affordable pricing.

According to Cahners In-Stat, the worldwide speech recognition technologies software market will increase to $2.7 billion by 2005. Datamonitor is even more optimistic, noting that the voice technologies market will grow to $5.6 billion by 2006. ASR technology has evolved to the point where it can easily be deployed at an enterprise level. Today, an average server can easily support applications that have vocabularies ranging in the thousands of words, with very high accuracy rates. As a DTMF replacement, speech recognition offers many improvements. It removes the limit as to the number of choices a user can have at a single time (DTMF is limited to zero through nine, star and pound). Rather than being limited to entering only numbers and letters, today's speech recognition systems can understand words and even phrases.

Text-to-speech, while not as popular as speech recognition, is also becoming more common. If you've tried calling information for a phone number, or calling to hear the latest weather or traffic conditions, you've probably experienced text-to-speech technology. The technology has come a very long ways from the early days when it was artificial sounding and difficult to understand. The new natural-sounding speech output is making new applications possible – such as being able to listen to your email, calendar and tasks over the phone. Today's text-to-speech systems can communicate information to callers that is either dynamically changing or from extremely large databases in a cost-effective and timely manner that would not be possible otherwise. One of the key reasons that TTS is growing in popularity is that it has a potentially infinite vocabulary, making it quick and easy to add new phrases without hiring actors and studios and having to record new prompts. TTS lets you use a standard voice throughout the system, whereas when using voice prompt-based systems you run the risk that the same actor may not be available when you want more prompts recorded, leading to an inconsistent user interface.

While speech technologies have greatly improved, customer expectations have increased as well. Users of these systems expect them to be easy to use, accurate, intuitive, and time efficient. While the user or customer only comes in contact with the speech-enabled application and user interface itself, the underlying technology of the speech system, or the platform, is of paramount importance. Key to a successful speech system is the communication adapter and software development kit (SDK) that system integrators, ISVs, and VARs can utilize in order to develop systems that provide users with a positive experience.

The Communication Adapter is Key

A communication adapter offers a variety of functions that greatly impact the overall performance of a speech-enabled system. The communication adapter is a key component for media processing, though not visible to the customer, and the choice of the right adapter makes a significant difference in the overall performance of system.

Different segments of users have different needs for speech-enabled applications, and these applications are essentially offered in two different ways. Telcos and other types of service providers can host the system and provide the speech applications, or enterprises can own and manage their own systems. Many enterprises outsource the development and management of speech applications to companies such as application service providers (ASPs) that specialize in speech technology. While there may be some different functions that enterprise and service provider users need, the bottom line is the same – all users demand accuracy, speed, and ease of use. Unless a speech-enabled system provides these capabilities, the end user experience will be unsatisfactory, resulting in lost customers and revenue for the company. ISVs, VARs, and system integrators can ensure that the speech system provides the functionality required by selecting the right communication adapter for their system.

What To Look For

There are several things ISVs, VARs, and system integrators should look for in a communication adapter. For speech-enabled systems, the communication adapter must have a key set of functions in order to allow developers to create a media processing system that results in a satisfying user experience. Without these components and capabilities, the speech system will not provide the high performance necessary to meet the challenges of demanding end-user customers, and companies will not experience the time and cost savings that can be realized by implementing speech systems.

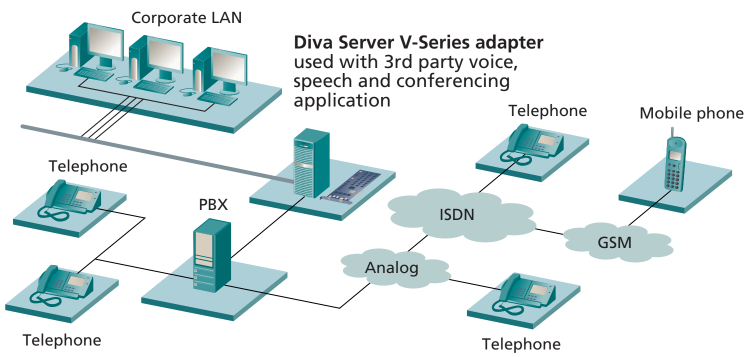
The most essential element is integration with leading speech engines currently on the market, which relies on three key capabilities – echo cancellation, full-duplex voice channel support, and voice activity detection (VAD). These capabilities in turn are necessary for one of the most important capabilities of a speech system – barge in, which allows the caller to interrupt the speech prompts by speaking over them. TTS and ASR can only really be used effectively when full-duplex, echo cancellation and barge-in are all supported.

As users, we've all experienced speech-enabled applications that make us listen to the entire menu before accepting our voice entries. This is very frustrating and leads to call abandonment by callers who do not want to waste time traversing long menus with a series of options to choose from. A well-designed speech application should allow barge-in, enabling the user to speak new commands even as the application is delivering information. Barge in is an essential element of any speech system. A basic requirement that is needed for barge-in is speech recognition in parallel with prompt playing (sometimes known as full duplex, or continuous speech processing). Full-duplex voice channels support is also needed to control the pace of the conversation and to help complete the interaction more quickly. The result is a more pleasant user experience and more efficient platform utilization.

Echo cancellation is an essential element for improved recognition accuracy, as well as for barge-in capabilities. Without echo cancellation, the speech recognizer gets confused when it "hears" its own outgoing prompts. In addition, due to the basic nature of networks, when users dial in over a public or private network there is often some type of echo on the signal that arrives at the server hosting the speech system. The better the system is at removing the echo or additional signal that is being reflected in the network, the more accurate the speech recognition.

Another major element, voice activity detection, reduces the volume of the voice stream and offloads host platform resources for speech recognition. By being able to detect a live voice and keep out extraneous noises, VAD can keep the line cleaner and only pass on real voices to the application. This not only saves bandwidth, but also makes it easier for the application to work effectively. While this function can be done in the PC, it is important to also have a board that can perform VAD.

Some speech recognizers cannot process speech and recognize DTMF at the same time, so using a voice board that can process DTMF on its own is another important requirement. There are several benefits to having the board process the DTMF, as this is a CPU-intensive task that is better accomplished via the board rather than the PC.



Benefits of Eicon's Diva Server V-Series in Speech Applications

When selecting a telephony board for speech-enabled systems, there are several considerations developers must take into account. Certainly, the key components discussed – barge in, echo cancellation, and full duplex voice channel support, are basic requirements. Telephony boards vary in the ways in which they provide these components, and selecting the right board is essential. Eicon Networks' Diva Server V-Series is a dedicated range of fully-functional Diva Server adapters that provide digital network interfaces and rich media processing for voice, speech, and conferencing. The Diva voice boards support speech technologies such as Speech Recognition, Voice Authentication, and Text-to-Speech, enabling developers to create a range of speech-enabled applications.

The Diva Server V-Series offers barge-in capabilities, full-duplex voice channel support, echo cancellation, and voice activity detection, with dedicated hardware resources for each channel, rather than having a pool of resources. Providing dedicated resources results in reduced application development time and lower costs due to the use of fewer boards.

Resource handling is done by Diva Server adapter, which simplifies the speech application development. Each DSP provides "universal" resources, such as voice, DTMF detection, tone detection, VAD, and conferencing, dedicated per channel. In contrast, boards from other vendors generally share resources of different types dynamically over channels, which is not as efficient. By having dedicated resources for the various channels, the telephony board, rather than the developer, manages and allocates the resources. The application does not need to make sure the channel is connected to the right type and number of resources, which is not the case when using shared resources.

In order to develop a speech recognition application, full-duplex voice (which requires two voice resources – one per stream of voice in each direction) and tone detection are both needed, thus requiring at least three resources per channel. Add echo cancellation and voice activity detection, and this increases to five resources. Without dedicated resources, an E1 board with 30 channels and 30 resources, which might require five resources per channel, would allow for only six concurrent calls before using up the resources on the board. This results in the need for additional boards, increasing the cost and complexity of the system.

By using a communication adapter with dedicated resources, the developer does not have to be aware of how the board works, what resources are available, or how to allocate the resources. For example, when barge-in is required, the board can automatically provide the resources needed without the developer having to take any special action to achieve this function, thus reducing overhead and development time. The communication adapter, not the script writer or developer, is responsible for managing the resources. Rather than requiring the developer to allocate the resources, the Diva Server V-Series board does this automatically. For example, when the call is complete, the Diva board automatically disconnects the resources. When using boards without dedicated resources, the developer would need to instruct the system when to disconnect the resource. With dedicated resources, the script writer or developer does not have to understand the architecture of the board in order to write the speech program. This is not the case with other types of boards that use pooled, rather than dedicated resources.

When running multiple types of media processing applications, such as speech recognition and text-to-speech, on a single server, the Diva Server V-Series can dynamically switch to the appropriate application without requiring the developer to write additional code. When a voice call comes in to the system, the communication adapter dynamically allocates the resources for echo cancellation, full duplex, and the other necessary components. The developer does not need to know what type of application or code is running on the board, as this is done in the background on the Diva Server. Alternatively, when using other types of telephony boards, the developer needs to write the extra code and allocate separate sets of resources for each type of application, resulting in additional time, expense, and overhead.

Exception handling is a specific area where using a Diva Server V-Series communication adapter with dedicated resources has tremendous advantages. When writing a speech application, approximately 80% of the code deals with exception handling, which is needed to deal with error conditions in the application. Since the exception handling and resource control is already done in the Diva Server V-Series board, the developer does not have to identify how to do exception handling and include it in the application code. By making the exception handling aspect of the code easier, the Diva Server greatly reduces the time and effort needed when developing voice applications.

Going Beyond the Basics

As the world moves to IP and voice over IP technology (VoIP) becomes more mainstream, existing voice applications will need to interwork with VoIP. Diva Server V-Series provides key enabling features for VoIP, such as voice packetization into Real-time Transport Protocol (RTP), voice compression (G.726 and GSM), Adaptive Jitter Buffer, and Comfort Noise Generation.

The speech recognition market has been moving toward standards, with two primary languages emerging - Voice Extensible Markup Language (VoiceXML or VXML) and Speech Application Language Tags (SALT). VXML is a system for developing and running standardized speech applications, and enables distributed applications by building on open Internet standards. Based on HTML, VXML is an open standard for telephony access and control that makes it easy to understand and program.

SALT is a lightweight set of extensions to existing markup languages, in particular HTML and XHTML, that enable multimodal and telephony access to information, applications and Web services from PCs, telephones, tablet PCs and wireless personal digital assistants (PDAs). While VXML and SALT are considered a layer above the Diva communication adapter, the Diva Server does support these standards and may be used for applications based on these languages.

There are also a range of supplementary services that may be used in speech-enhanced applications. Call Transfer functions can be used, for example, in call routing or soft-PBX-type functions. Call Transfer is specified as part of VXML. While some telephony boards do not hide the complexity of this from their users (in some cases the users are effectively asked to build their own signaling messages and drop them onto the D channel), Eicon's call transfer (and explicit call transfer) options are implemented in a way that makes it easy to perform these functions from an SDK application.

Diva Server V-Series also provides integration with ASR and TTS speech engines, and provides the drivers to make it easier for application partners to work with the speech engines. In addition, the Diva Server SDK, a software development kit, can be used to simplify Diva Server integration in speech applications by providing media streams directly to speech engines. It also provides tools, utilities and samples, and application programming interfaces (APIs), allowing software vendors to easily develop their own communication applications for all Diva Server V-Series adapters.

Eicon's Diva Server V-Series – Optimized for Speech

In order to build a speech-enabled solution, ISVs and system integrators may choose from a variety of hardware platforms. Eicon Networks' Diva Server V-Series is a family of intelligent telephony boards optimized for voice and speech applications, based on Eicon's established and popular Diva Server range of intelligent ISDN adapters.

The Diva Server V-Series family ranges from a single Basic Rate (BRI) board to multi-BRI as well as Primary Rate (PRI), E1 and T1 models. Designed to be scalable, any combination of Diva Server V-Series boards can be mixed and matched in a system. As opposed to other vendors offering a complex range of boards, components and expansions that do not seamlessly fit together, Eicon's communication adapters all use a single set of driver software and a single set of APIs, enabling ISVs to develop applications that work across the entire range of Eicon products. All of Eicon's boards, whether BRI, T1, E1, or any other, share the same architecture of RISC CPU and Digital Signal Processors, allowing developers to mix and match boards in a single machine – the application stays the same regardless of which Eicon board is being used.

Diva Server V-Series support a range of interfaces, from base rate ISDN to primary rate, offering 2, 8, 24, and 30 channels of capacity under this family of boards. Network interfaces range from a single analog port up to multiple PRI/E1/T1, while providing superior scalability from two to 240 channels. The onboard DSP resources provide a full set of voice processing functions, including tone detection and generation, DTMF handling, audio tapping, voice activity detection and echo cancellation.

Conclusion

The desire for profitable and satisfactory customer relationships is stronger than ever, and the role of speech technologies to help provide exceptional service has never been greater. The use of speech technologies is enabling companies of all sizes to better interact with customers more profitably by increasing self-service capabilities, while increasing customer satisfaction. Speech-enabled applications will continue to dramatically increase for the foreseeable future, and application developers, system integrators, and VARs will be leading the way with new and innovative systems. However, choosing the right communication adapter for media processing is crucial.

Eicon Networks' Diva Server V-Series provides a full set of voice processing functions, as well as efficient integration with speech engines. The powerful hardware architecture and rich media processing features of Diva Server V-Series boards allow companies to develop and deploy speech applications that deliver enhanced performance, greater accuracy and reduced costs. By developing speech systems based on the Diva Server V-Series, developers and integrators can reduce their time to market and total cost of ownership, while providing the capabilities and functionality that their customers need.

Corporate headquarters

Eicon Networks Corporation

9800 Cavendish Blvd 5th Floor
Montreal Quebec Canada
H4M 2V9
Tel: +1 (514) 745-5500
Fax: +1 (514) 745-5588

Regional sales head offices

Americas:

Eicon Networks Inc.

Parkway Centre II
2805 N. Dallas Parkway Suite 200
Plano Texas 75093 USA
Tel: +1 (972) 473-4500
Fax: +1 (972) 473-4510

Offices: Dallas, Montreal

Europe, Middle East, Africa:

Eicon Networks (UK) Ltd.

Kings Chase 107-123 King Street
Maidenhead Berkshire SL6 1DP
United Kingdom
Tel: +44 (0) 1628 641770
Fax: +44 (0) 1628 641780

Offices: Barcelona, Bergamo, Stockholm, Den Haag, Ljubljana, London, Paris, Leonberg, Berlin, Dusseldorf, Munich

Asia Pacific:

Offices: Beijing, Hong Kong, Kuala Lumpur, Shanghai, Sydney

© 2004 Eicon Networks Corporation. All rights reserved. All trademarks or registered trademarks are the property of their respective owners. [GR04/13/084] 221-539-01