



# **Performance Characterizations for Speech Applications: Microsoft Speech Server\* and Intel® Telephony Servers**

## Contents

---

<b>Executive Summary</b>	<b>1</b>
<b>Introduction</b>	<b>1</b>
<b>Speech-Enabled Architecture</b>	<b>2</b>
<b>Test Configuration</b>	<b>3</b>
<b>Results</b>	<b>4</b>
<b>Conclusions</b>	<b>5</b>
<b>Product List</b>	<b>5</b>
<b>Definitions and Acronyms</b>	<b>5</b>
<b>For More Information</b>	<b>6</b>

---

## Executive Summary

This application note describes test results that demonstrate the scalability of Microsoft Speech Server\* 2004 Enterprise Edition to a level capable of supporting large-scale enterprise deployments.

Microsoft Speech Server's scalability was tested in a multi-server configuration incorporating Intel® Pentium® processors, Intel® telephony boards, and Intel NetMerge® Call Manager software. The results provided here show that a 1,100-port configuration with a call completion rate of more than 1,700 calls per minute experienced an aggregate CPU utilization rate under 50%.

## Introduction

Improvements in speech technology and the increased use of PDAs, mobile voice, and multi-modal devices are combining to propel the development and deployment of speech-enabled applications. Incorporating speech as a viable user interface offers a degree of flexibility to users who previously have been limited to fix-placed Web browsers or hierarchical Interactive voice response (IVR) menu schemes.

Automatic speech recognition (ASR) and text-to-speech (TTS) are two of the baseline services that make it possible to extend application user interfaces beyond Web browser and touchtone (DTMF) interactions. Speech technology is further enhanced by the expanding use of standards such as Speech Application Language Tags (SALT), which establishes a common markup language for carrying interactive speech between devices that are so enabled. Intel refers to the convergence of speech and established Web interface technologies as Communications Web Services (CWS).

## Challenges

There are three specific challenges to adding speech and evolving to a CWS architecture:

1. Finding speech tools that are compatible with current development environments
2. Supporting new speech-capable devices as well as traditional wireline and wireless devices
3. Deploying speech technology without significantly impacting system resource utilization

## Reference Architecture

To meet these needs, Intel and Microsoft have developed a reference architecture that includes the following components:

- Microsoft Speech Server\* 2004
- Intel® NetStructure™ combined media boards
- Intel NetMerge® Call Manager software
- Intel® Dialogic® System Release software

Scalability performance was tested in configurations that incorporated all of these components.

Speech system performance depends on several factors including:

- The vocabulary involved
- CPU speed
- Simultaneous channel count
- System architecture

The test results presented in this application note are representative. You should use them as a general guide, and not as the basis for planning a specific deployment.

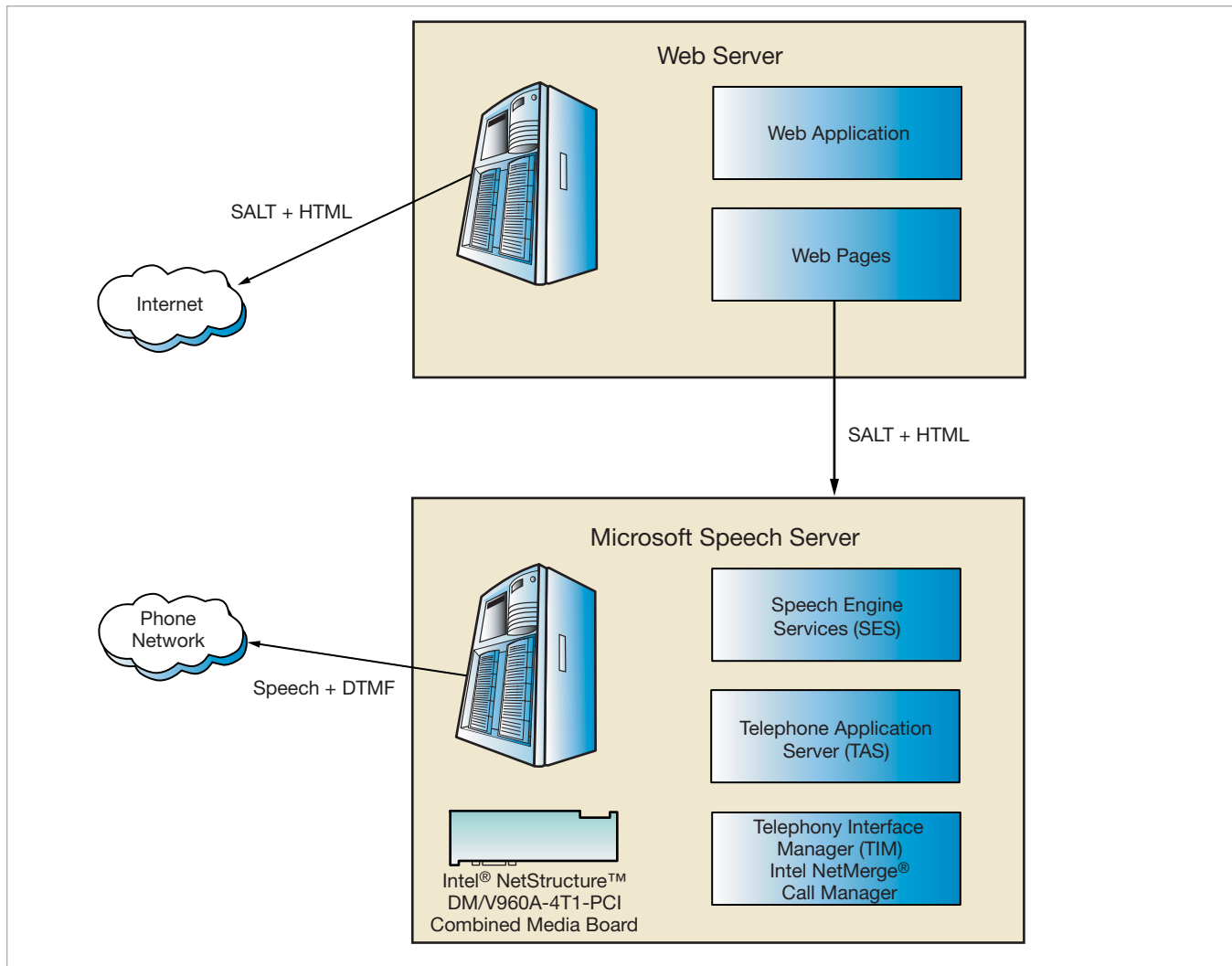


Figure 1: Speech-Enabled Architecture

## Speech-Enabled Architecture

An application Web server that generates SALT-embedded HTML pages is capable of communicating directly with multi-modal devices such as PCs or smartphones. But communicating with end user devices that support only traditional audio (e.g., telephone handsets) requires a means to translate SALT and content text into synthesized or recorded speech. Input from user devices requires translation on the server side from voice responses and DTMF tones into a format recognized by the application.

Microsoft Speech Server is a platform that extends ASP.NET Web applications by providing a translation point between the Web server and audio or speech-enabled devices without impacting existing Web application architectures. To customize an application, software designers use the Microsoft Speech Application Software Development Kit\* (SASDK\*) to create the grammars and interaction rules that generate the speech

interface. SASDK is compatible with Microsoft Visual Studio\* .NET 2003.

As shown in Figure 1, Microsoft Speech Server consists of two high-level services that provide the speech-enabling translation: Speech Engine Services (SES) and Telephony Applications Services (TAS).

SES manages grammars and voice prompts and provides the foundation ASR and TTS services necessary to implement server-side speech functionality. These speech resources isolate the Web server from differences in the capabilities of end user devices.

TAS facilitates the SALT-to-audio process through the instantiation of two sub-layers:

- Core services
- Telephony Interface Manager (TIM)

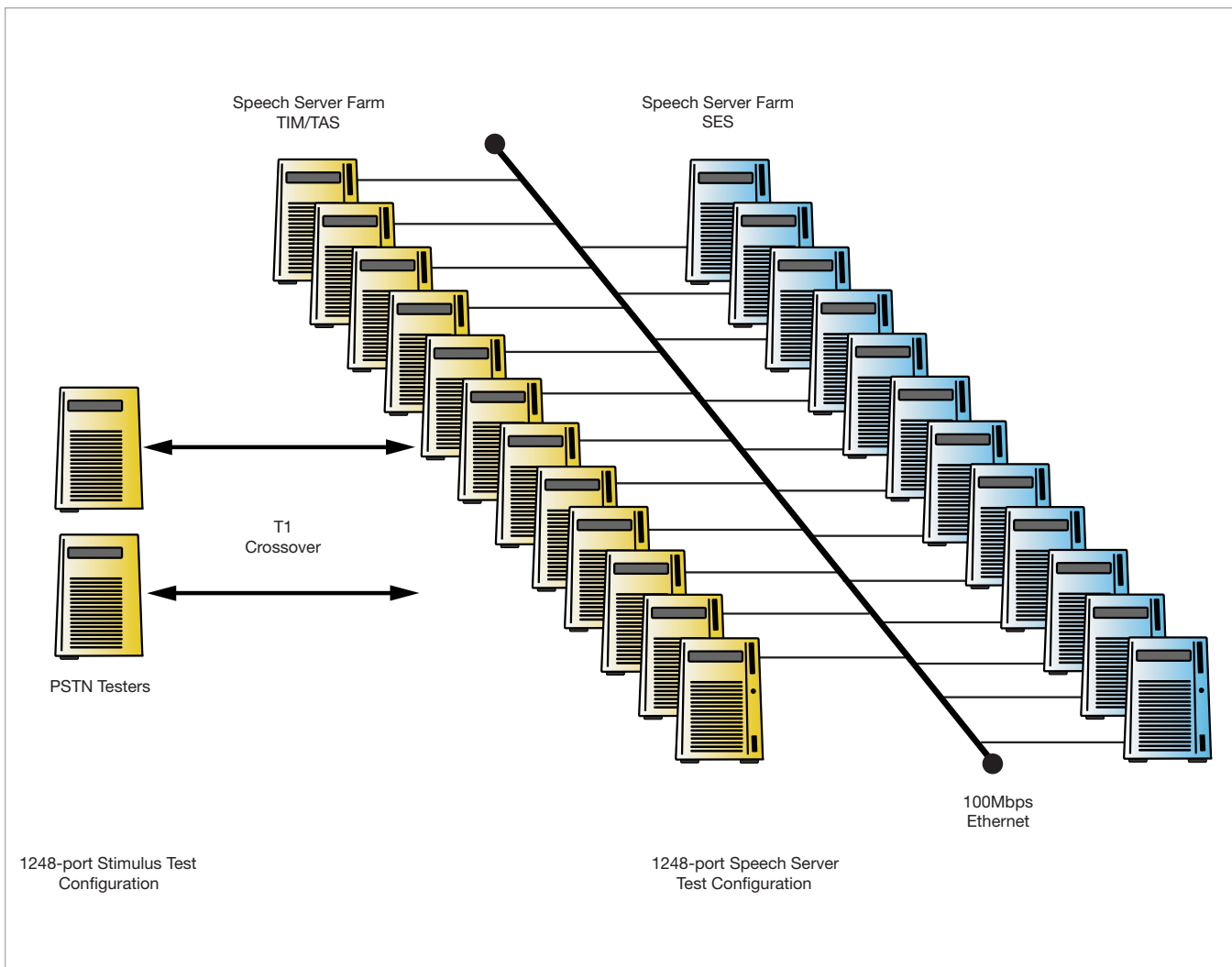


Figure 2: Test System Configuration

The core services include SALT interpretation, call control, call logging, and DTMF translation. TAS correlates the state and session information between the Web server and the established calls.

TIM is a software layer that links the core services with specific telephony boards. It carries out control functions such as establishing and terminating calls. By contrast, the core services operate at a higher level of abstraction and remain the same across all telephony environments.

Intel NetMerge Call Manager software is Intel's implementation of the TIM. It communicates with the physical telephony hardware, abstracting the SES and TAS core services from telephony resources. It handles all call control functionality (e.g., transferCall, makeCall, answerCall) and media streaming to and from the SES. Call Manager supports a variety of telephony boards, enabling a wide range of high- and low-density digital and analog applications.

Each TAS/TIM test system included the Intel NetStructure DM/V960A-4T1-PCI combined media board, which provides 96 channels of T-1 voice along with voice processing and conferencing. The board supports Intel® Dialogic® Continuous Speech Processing Technology, an advanced speech pre-processing technology that scans incoming audio signals and removes silence and background noise from valid speech patterns. This feature significantly reduces the CPU requirements for the TAS/TIM by delivering only the essential voice payload to the host.

## Test Configuration

The test configuration was based on a distributed architecture as illustrated in Figure 2. Each test call was approximately 20 seconds in duration and scripted as follows:

- The stimulus system placed a call to the system under test (SUT). Each call was routed to the first available port.

## Application Note

### Performance Characterizations for Speech Applications: Microsoft Speech Server\* and Intel® Telephony Servers

- Upon answering each call, the SUT played a prompt via TTS and awaited voice input.
- The stimulus system played one of several available voice prompts in reply.
- In response to the specific voice prompt, the test system responded with a corresponding DTMF identifier.

Each server was configured with an Intel® Server Chassis SR2300, dual 2.8 GHz Intel® Pentium® 4 processors, and 4 Gigabytes of RAM, running Microsoft Windows\* Server 2003 Enterprise Edition and the following additional components:

- **SES server**
  - Microsoft Speech Server 2004 Enterprise Edition SES Component
- **TAS/TIM server**
  - Microsoft Speech Server 2004 Enterprise Edition TAS Component
  - One (1) Intel NetStructure DM/V960A-4T1-PCI combined media board
  - Intel NetMerge Call Manager software
  - Intel® Dialogic® System Release 6.0 PCI for Windows\*
- **Test stimulus server**
  - Six (6) Intel NetStructure DM/V960A-4T1-PCI combined media boards
  - Intel NetMerge® Media Communications Carrier Software to process the outbound calls to the TAS/TIM servers
- **Opus75 PNT (Public Network Tester)**

**Note:** SES and TAS/TIM were run on separate servers as required by Microsoft Speech Server Enterprise Edition. A production system running Microsoft Speech Server Standard Edition may run both components on the same server, with consideration given to server size, port density, and application resource requirements.

## Results

The test results are shown in Table 1 for both 460-port and 1,104-port configurations.

The aggregate per-server CPU utilization for the 460-port configuration was 29%, while the utilization for the 1,104-port configuration was 47%. The relatively flat utilization for the TAS/TIM server is accounted for by the activation of the Continuous Speech Processing Technology in the Intel NetStructure DM/V960A combined media board, which can decrease utilization by 20% to 37% because it passes only valid speech data to the host.

The call completion rate for a 460-port configuration was 1,091 calls per minute, or 2.4 calls per minute per port. The rate was 2,023 calls per minute for the 1,104-port configuration, or 1.6 calls per minute per port.

The lower call rate for the larger configuration is largely accounted for by the use of a first available port selection strategy. Microsoft Speech Server is capable of supporting a load balancing front end. It is likely that the average calls per port per minute rate will improve when load balancing hardware or software is used.

**Table 1: Test Results**

Port Count	Test Duration (Hours)	Average Calls/Minute	Average Calls/Port Minute	Completion Rate (%)	SES CPU (%)	TAS/TIM CPU (%)
460	93	1,091	2.4	>99	11	18
1,104	71	1,756	1.6	>99	32	15

## Conclusions

CWS platforms have the power to enable rapid integration of speech into existing or new Web-based applications. Microsoft Speech Server is a scalable platform that extends the capabilities of ASP.NET Web applications to include speech as an equivalent, full-function user interface. When combined with telecommunication boards from Intel that include Intel Dialogic Continuous Speech Processing Technology, Microsoft Speech Server can scale from small or medium offices up to large deployments supporting contact center or unified communications solutions without affecting system resource utilization.

Individual system performance will vary significantly according to application resource utilization, network configuration, and other variables. However, these test results and test configurations should supply a baseline of performance and density that can provide a foundation for successful speech deployments.

## Product List

### Hardware

Intel® Server Chassis SR2300 with dual 2.8 GHz Intel® Pentium® 4 processors

Intel® NetStructure™ DM/V960A-4T1-PCI Combined Media Board

### Application Software

Microsoft Windows Server 2003 Enterprise Edition

Microsoft Speech Server 2004 Enterprise Edition

Intel NetMerge® Call Manager

Intel® Dialogic® System Release 6.0 PCI for Windows\*

Intel® Dialogic® Global Call Development Software

### Test Software

Intel NetMerge® Media Communications Carrier Software

Opus75 PNT (Public Network Tester) OpenCLI

## Definitions and Acronyms

<b>ASR</b>	Automatic speech recognition
<b>CWS</b>	Communications Web Services
<b>DTMF</b>	Dual tone multi-frequency
<b>HTML</b>	HyperText Markup Language
<b>IVR</b>	Interactive voice response
<b>PDA</b>	Personal digital assistant
<b>SALT</b>	Speech Application Language Tags
<b>SASDK</b>	Microsoft Speech Application Software Development Kit
<b>SES</b>	Speech Engine Services
<b>SUT</b>	System under test
<b>TAS</b>	Telephony Applications Services
<b>TIM</b>	Telephony Interface Manager
<b>TTS</b>	Text-to-speech

## For More Information

### Articles

- Article: Microsoft Speech Technologies Telephony Server Platform — [http://www.microsoft.com/speech/docs/Telephony\\_Server\\_Platform.htm](http://www.microsoft.com/speech/docs/Telephony_Server_Platform.htm)

### White Papers

- Intel White Paper: Communications Web Services: Bringing the Value of the Web to Voice Solutions — <http://www.intel.com/network/csp/pdf/8857wp.htm>
- Intel White Paper: Continuous Speech Processing: Speech Technology Enters the Mainstream — [http://www.intel.com/network/csp/resources/white\\_papers/6556web.htm](http://www.intel.com/network/csp/resources/white_papers/6556web.htm)
- Microsoft White Paper: An Introduction to Telephony Call Control with Microsoft Speech Server — <http://www.microsoft.com/speech/evaluation/whitepapers/TelephonyCallControlDoc/>
- Microsoft White Paper: Load Balancing Schemes with Microsoft Speech Server — <http://www.microsoft.com/speech/docs/WhitePaperLB.htm>

### Product Information

- Intel NetMerge Call Manager — <http://www.intel.com/network/csp/products/8661web.htm>
- Intel Dialogic Global Call Development Software — <http://www.intel.com/network/csp/products/2557web.htm>
- Intel Dialogic System Release 6.0 PCI for Windows — <http://www.intel.com/network/csp/products/8830web.htm>
- Intel NetStructure DM/V960A-4T1-PCI Combined Media Board — <http://www.intel.com/network/csp/products/7603web.htm>
- Intel Server Chassis SR2300 — [http://www.intel.com/design/servers/sr2300/?iid=ipp\\_svr+svrchassis\\_sr2300&](http://www.intel.com/design/servers/sr2300/?iid=ipp_svr+svrchassis_sr2300&)
- Microsoft Speech Server 2004 Enterprise Edition — <http://www.microsoft.com/speech/evaluation/>
- Microsoft Windows Server 2003 Enterprise Edition — <http://www.microsoft.com/windowsserver2003/evaluation/overview/family.mspx>

To learn more, visit our site on the World Wide Web at <http://www.intel.com>.

1515 Route Ten  
Parsippany, NJ 07054  
Phone: 1-973-993-3000

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

Intel, Intel NetMerge, Pentium, Intel Centrino, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

\*Other names and brands may be claimed as the property of others.

