



# Speech Recognition

in Version 2.3 of the *Interaction Center Platform*<sup>®</sup>

# INTERACTIVE INTELLIGENCE

## Overview

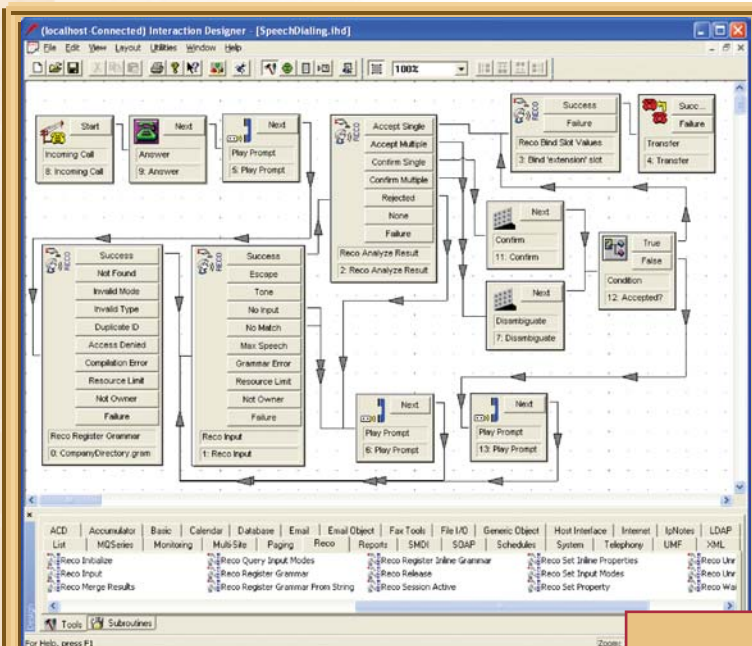
After years of unfulfilled promises, speech recognition is now moving into the mainstream. The technology has progressed to the point that it is now possible to reliably recognize specific words (such as "savings") and even whole sentences (such as "I want to buy one thousand shares of IBM at market") when spoken over the phone by people with a broad range of accents. Organizations that have deployed speech-enabled customer service applications also have realized tremendous cost savings, and have sometimes been able to work around thorny patent issues associated with traditional interactive voice response (IVR).

Unfortunately, accessing this newly matured capability has been difficult. You've needed C++ or Java programmers, grammar designers and sophisticated development tools to create meaningful speech-enabled applications. And maintaining and modifying such applications has meant dealing with multiple servers, engines, and other components.

## The Interactive Intelligence Approach

The Interaction Center Platform from Interactive Intelligence<sup>®</sup> makes powerful speech recognition capabilities more accessible than ever before. In other words, it wraps engines from leading vendors such as ScanSoft<sup>®</sup> (the SpeechWorks<sup>®</sup> engine), Nuance, and Aculab with an "engine-agnostic" interface that allows applications to be constructed from high-level building blocks that are simpler to put together and easier to maintain.

What's more, the speech recognition grammars used to describe what callers are expected to say are specified using an industry standard format, eliminating another source of dependency on specific vendors. Over time, the Interaction Center Platform will provide built-in speech-enabled applications for auto attendant, voice mail retrieval, contact management, and many others. In addition, customers can use the Platform's inherent *Interaction Designer*<sup>®</sup> graphical application generator to create custom applications that blend speech recognition, database access, telephony, and just about anything else you can imagine.



Built-in Applications

Custom Applications

Engine-Agnostic Interface

SpeechWorks Engine

Nuance Engine

Aculab Engine

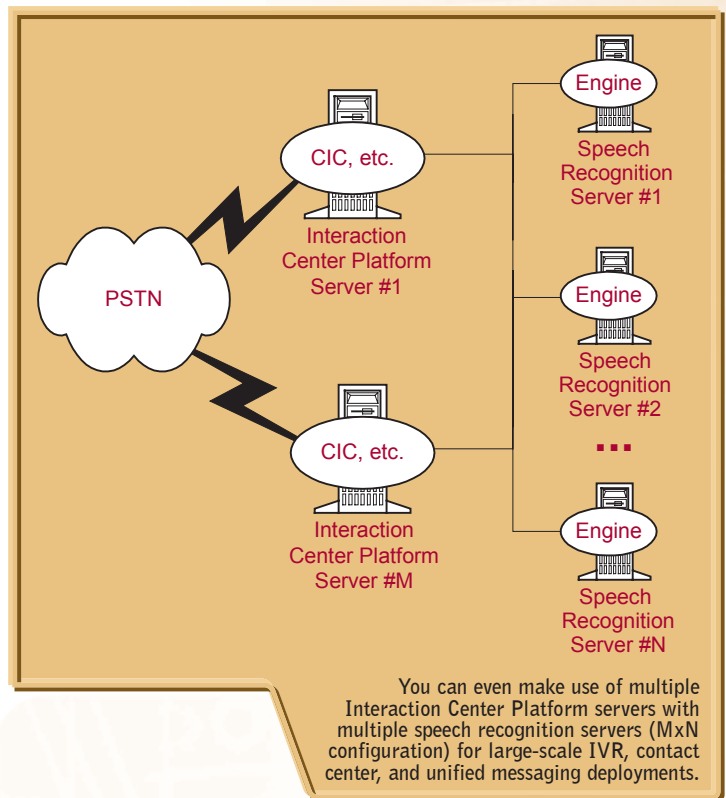
Other Engines

*Interaction Designer* includes numerous speech recognition building blocks that can be used to create sophisticated speech-enabled applications both for internal and external use

## Benefits and Advantages

The Interaction Center Platform breaks new ground in making speech recognition technology accessible. Compared to the speech recognition capabilities in other communications products, the Interaction Center Platform provides:

- **Unlimited scalability and high reliability.** Set up as many speech recognition servers as you like. The Interaction Center Platform performs automatic load-balancing across them, allowing you to handle thousands of simultaneous calls. If a server goes down, calls are immediately routed to another one. You can even make use of multiple Interaction Center Platform servers with multiple speech recognition servers (MxN configuration) for large-scale IVR, contact center, and unified messaging deployments.
- **Ease of administration.** Use the Platform's *Interaction Administrator*<sup>®</sup> graphical administrative console to configure one or many speech recognition servers.
- **Real-time supervision.** The Interaction Center Platform's real-time *Interaction Supervisor*<sup>™</sup> supervisory console allows you to monitor all your speech recognition servers and sessions.
- **Your choice of engines.** Avoid being locked into a particular speech recognition engine. The Interaction Center allows you to make use of engines from leading vendors including Nuance, ScanSoft (SpeechWorks), and Aculab.
- **Simultaneous support of different engines.** Each speech recognition engine has its own unique advantages and disadvantages. The Interaction Center Platform allows you to use different engines (e.g., SpeechWorks and Aculab) at the same time. For example, an application requiring complex grammar support could be serviced by a high-end engine, while a simple command-driven voice interface could be handled by a much less expensive low-end engine. This allows you to use the right engine for the right job and potentially save lots of money.
- **Adherence to industry standards.** The Interaction Center Platform also allows grammars to be specified in the SRGS (Speech Recognition Grammar Specification) format with semantic interpretation tags specified in the SISR (Semantic Interpretation for Speech Recognition) format. Both SRGS and SISR were developed by the World Wide Web Consortium (W3C), the standards body behind the Internet.



- **Rapid application development.** The Interaction Designer graphical application generator can be used to quickly create powerful speech recognition applications using a set of pre-built components. XML facilities allow you to dynamically create grammars and process results for highly personalized applications.
- **Multiple language support.** Speech recognition applications can simultaneously handle all the dozens of languages supported by the underlying engines—from Japanese to Hebrew. Easily route a call to the engine with the best support for the caller's language, or dedicate an entire server to handling a particular language.
- **Speaker verification.** Applications can use advanced speaker identification and verification technology to authenticate callers for sensitive transactions.
- **Unified results for speech and DTMF.** Avoid creating separate applications for speech and touch tones. The Interaction Center Platform allows you to easily handle either form of input and process the results identically.
- **Preloaded grammars.** Designated grammars can be automatically distributed to all speech recognition servers and pre-loaded for use. This is particularly useful for large, dynamically-generated grammars, such as a dial-by-name company directory.
- **RTP streaming.** Voice traffic is streamed over an IP network to the speech recognition servers using the standard RTP protocol used in voice over IP (VoIP) applications. This approach removes any dependency between the speech recognition engine and the telecommunications equipment. It also allows the speech recognition servers to be located remotely.



## Specifications

### Capacities

- Number of recognition sessions per server - varies with engine, application complexity, and CPU
- Number of speech recognition servers - unlimited
- Number of grammars used per call - unlimited

### General features

- Intelligent grammar caching
- Use of the SRGS grammar specification (both XML and ABNF syntaxes)
- Web-based administration of speech recognition servers
- Dynamic propagation of configuration changes
- Ability to handle both speech and DTMF input in the same application
- Grammars can be referenced as files, URLs, built-in names, or from text strings
- Built-in grammars for yes/no and digits for all engines
- Support for VoiceXML built-in grammars (date, time, currency, number, phone number, credit card number, social security number) for SpeechWorks and Nuance engines
- Supports semantic tag syntax in the SRGS grammars according to the SISR specification
- Recognition results returned as XML for easy and unified processing
- Multiple hypotheses (possible recognition matches) returned in the result with confidence values
- Fine-grained timeout control
- Lazy port allocation
- Barge-in support
- Automatic load-balancing across multiple speech recognition servers
- Extensive application development facilities for looping, disambiguation, XML processing, database access, etc. using Interaction Designer

### Application building blocks available in Interaction Designer

- **RECO Initialize** - Initializes a session for a call using a specific engine and language, if desired
- **RECO Release** - Releases a recognition session
- **RECO Session Active** - Checks to see if a recognition session is active on the current call
- **RECO Has Feature** - Checks to see if the engine servicing the current call has a particular feature (e.g. ability to support multiple grammars)
- **RECO Get Property** - Get the value of the specified property of the engine servicing the current call (e.g. barge-in mode)
- **RECO Set Property** - Set the value of the specified property of the engine servicing the current call
- **RECO Query Input Modes** - Checks to see if the current recognition session handles speech and/or DTMF input
- **RECO Set Input Modes** - Specifies whether the current recognition session should handle speech and/or DTMF input



- **RECO Input** - Uses one or more grammars to try to recognize input by speech and/or DTMF
- **RECO Basic Input** - Uses one of the built-in grammars to recognize simple input (e.g. "Yes" or "No")
- **RECO Register Grammar** - Registers a grammar for use in the current call
- **RECO Register Grammar String** - Compiles and registers a grammar from a text string
- **RECO Register Inline Grammar** - Register small inlined grammar associated with language of the call.
- **RECO Unregister Grammar** - Removes a grammar from use in the current call
- **RECO Unregister All Grammars** - Removes all currently registered grammars from use in the current call
- **RECO Get Registered Grammars** - Returns a list of currently registered grammars
- **RECO Wait For Grammars** - Bounded wait for asynchronously registered grammars to become ready for use.
- **RECO Add Preloaded Grammar** - Parses, compiles, and pre-loads a grammar into all engines for seamless changeover of very large grammars such as dial-by-name grammars.
- **RECO Get Hypothesis** - Returns a specified portion of a result in XML form from a recognition by RECO Input or RECO Basic Input
- **RECO Get Next Hypothesis** - Iterates over the portions of a recognition result
- **RECO Analyze Result** - Figures out whether a recognition result is reliable enough to accept or needs to be disambiguated
- **RECO Get Slot Value** - Given a recognition result and the name of a semantic slot (e.g. "City") returns the corresponding value (e.g. "Indianapolis")
- **RECO Bind Slot Values** - Given a recognition hypothesis, binds the content of multiple semantic slots to variables
- **RECO Merge Results** - Merges the result of multiple recognition events, useful in disambiguation
- **RECO Filter Result** - Returns the parts of a recognition result matching a given pattern
- **RECO Custom Operation** - Allows an application to access low-level features of a specific speech recognition engine

## Requirements

### Supported speech recognition engines

- Nuance 8.0
- SpeechWorks (ScanSoft) Open Speech Recognizer 1.1
- Aculab Connected Word Recognizer

### Server machines (where speech recognition engines are installed)

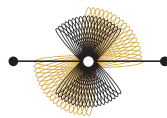
- Intel®-based systems configured according to the recommendations of the engine vendor
- Microsoft® Windows® 2000 or Windows® 2003 server operating system
- Must be dedicated to execution of the speech recognition engine and necessary Interactive Intelligence components
- Must NOT be the same server running an Interaction Center Platform product (e.g. *Customer Interaction Center®*)

### Supported telephony platforms

- Intel® Dialogic® - requires boards supporting CSP or ECP-capable voice resources (JCT trunk boards, DM/T or DM/V boards with media load 2 and 10, and DI/0408 (Denali media load 4))
- Aculab - requires version 1.5 and later Prosody boards with sufficient DSPs for echo cancellation
- SIP - with the supported Intel/Dialogic or Aculab boards
- Cisco AVVID (via TAPI) – planned for a feature pack scheduled for the first quarter of 2004

## For More Information and Availability

For more information on the *Interaction Center Platform* technology, speech recognition, and our suite of IP telephony-based interaction management software products, visit [www.ININ.com](http://www.ININ.com) or e-mail [info@ININ.com](mailto:info@ININ.com).



INTERACTIVE INTELLIGENCE®

#### WORLD HEADQUARTERS

7601 Interactive Way  
Indianapolis, IN 46278 USA  
317.872.3000 voice and fax

#### EUROPE

8 The Square, Stockley Park  
Heathrow, Uxbridge, Middlesex, UB11 1FW  
+44 (0) 20.8610.6045 voice; +44 (0) 20.8610.6095 fax

#### ASIA PACIFIC

Suite 24.3, Level 24 Menara IMC  
8 Jalan Sultan Ismail, 50250 Kuala Lumpur, Malaysia  
+603.2715 3333 voice; +603.2715 3343 fax

[www.ININ.com](http://www.ININ.com)