Voice Processing Standards

Mukesh Sundaram
Vice President, Engineering
Genesys (an Alcatel company)

Agenda

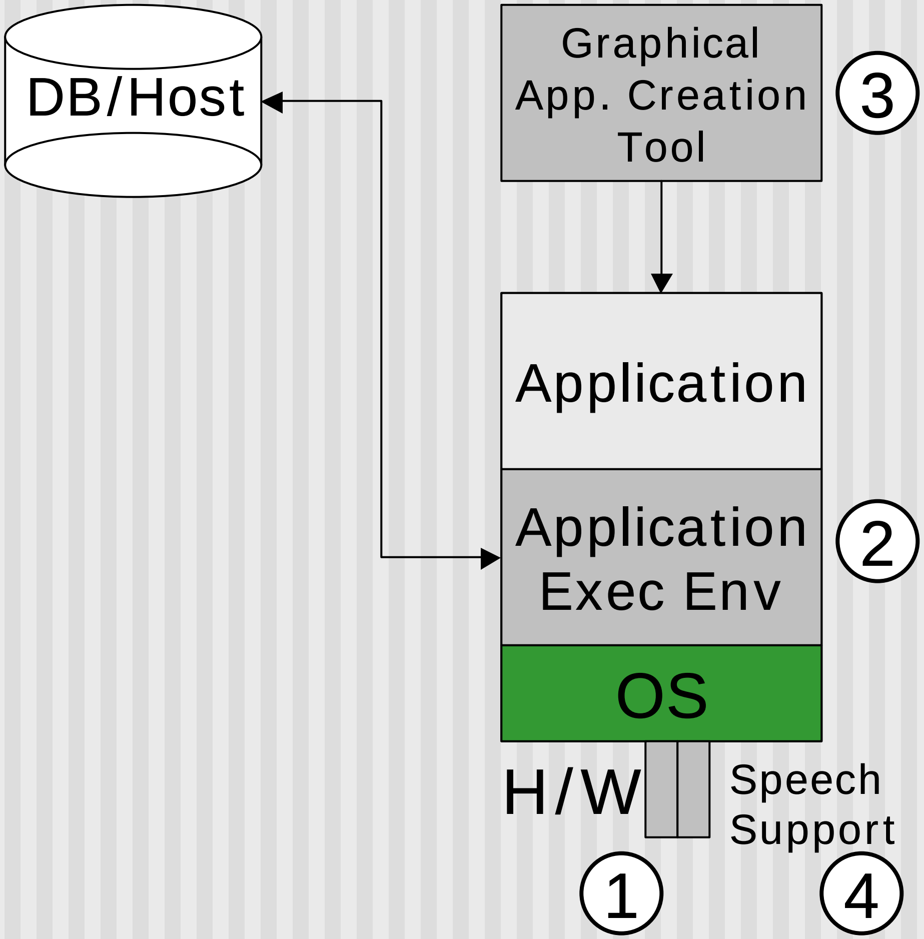
- Interactive Voice Response
- Speech Processing
- Computer Telephony Integration
- IP Telephony
- Standards Activities
- Impact on Industry and Customers
- Outlook

Interactive Voice Response

The IVR System

- Interactive Voice Response System
- Replaces human operators
- Evolved from proprietary roots
  - Special application & hardware
- Commercialized as integrated hardware and software product from vendors
- DTMF based systems could not be upgraded for Automatic Speech Recognition (ASR)
  - Forklift hardware upgrade
- One of the more hated system in the Enterprise
  - Vendor provided everything—hardware, software, and application development tools

Parts of an IVR



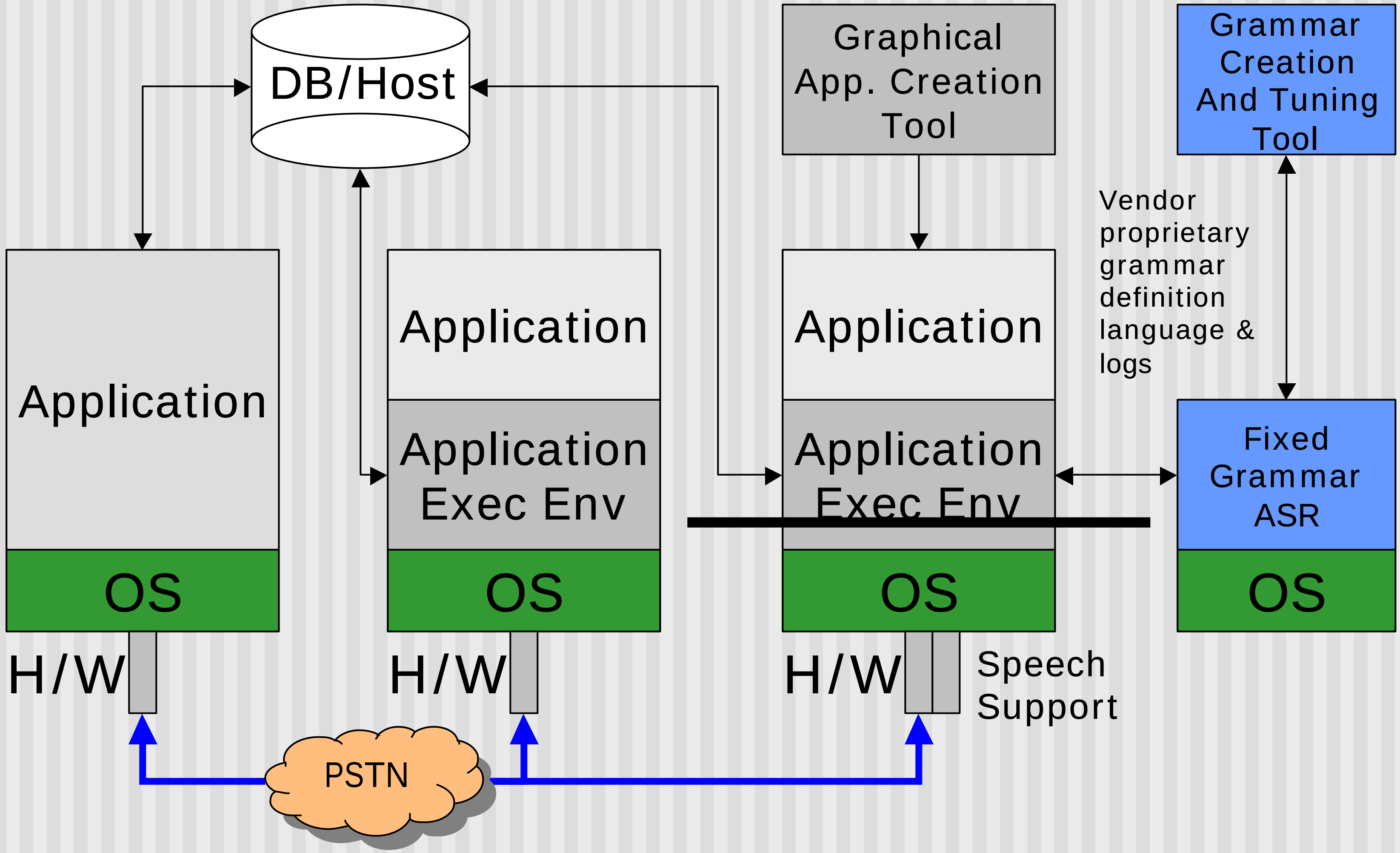
1. Proprietary hardware+OS
2. Proprietary application execution environment and hooks to databases/legacy hosts
3. Proprietary GUI based application creation tool
4. Proprietary speech support hardware

IVR Evolution and Roadmap

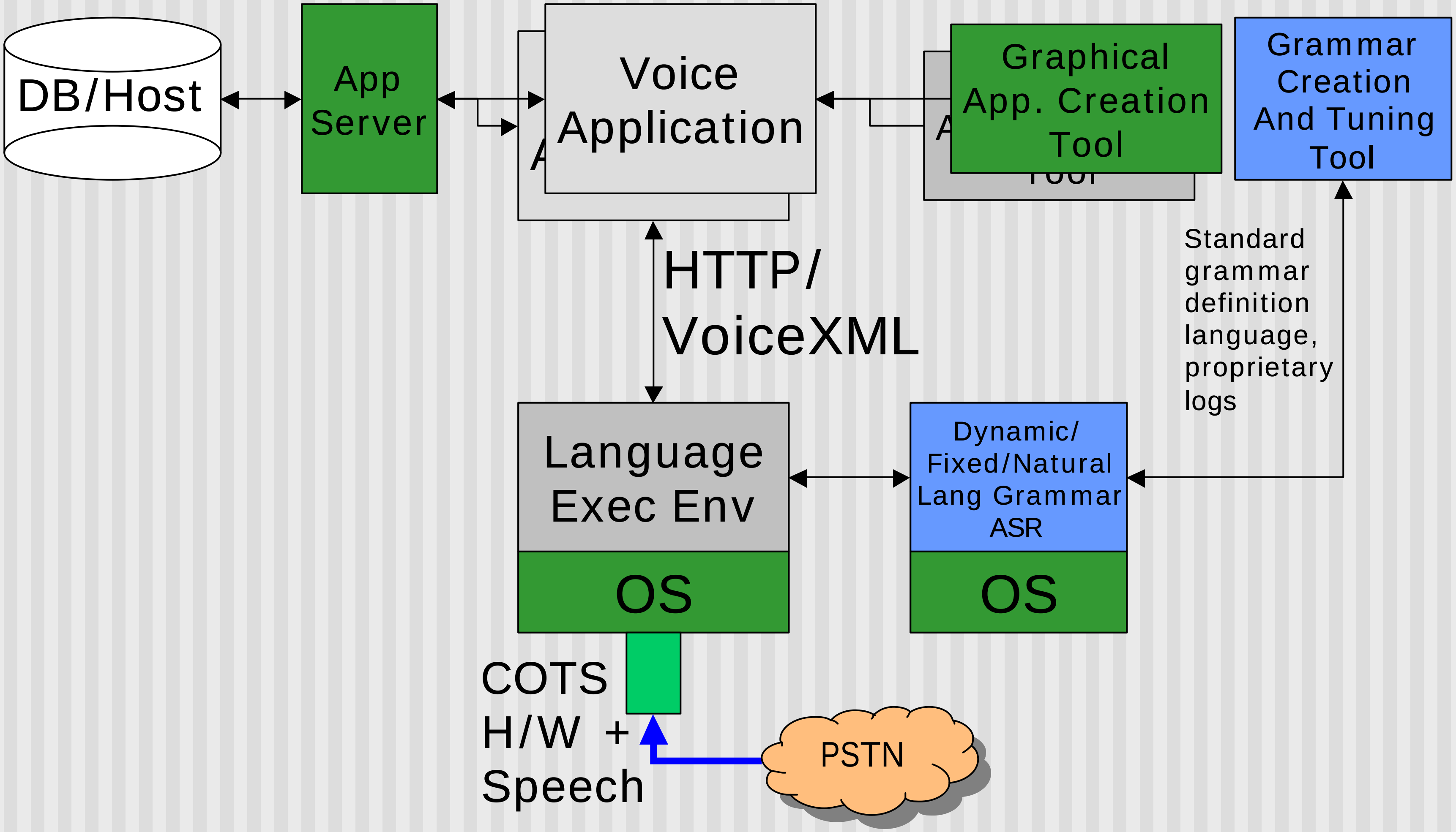
Market Place

	Phase I	Phase II	Phase III	Phase IV
Hardware, OS, Telecom	**Proprietary**	**Proprietary**	**Generic HW/OS, Telecom board**	**Server farm, Pure software, Full IP**
Application development	**Proprietary**	**Proprietary, graphical**	**Internet technology, tools**	**Web & Voice integrated tools**
Back end integration	**None**	**Ad hoc**	**Internet standards based**	**Web & Voice part of back end framework**
User Interface	**Menu, DTMF**	**Menu, “say yes”**	**Mature ASR/TTS application (VUI)**	**Natural Language**

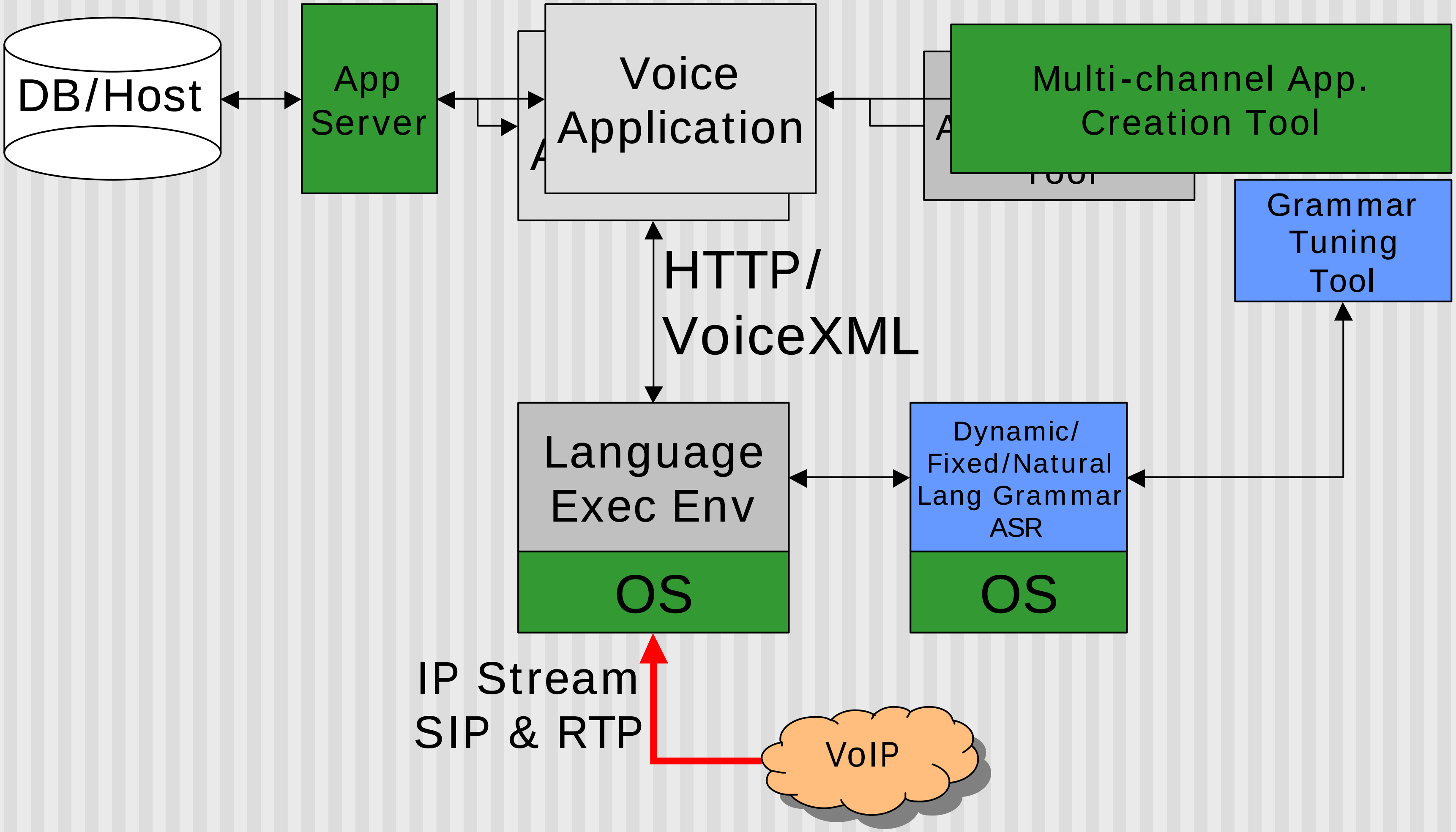
IVR Systems: Phase I/II



IVR Systems: Phase III


DB/Host
App Server
Voice Application
Graphical App. Creation Tool
Grammar Creation And Tuning Tool
HTTP/ VoiceXML
Standard grammar definition language, proprietary logs
Language Exec Env
OS
Dynamic/ Fixed/Natural Lang Grammar ASR
OS
COTS H/W + Speech
PSTN

IVR Systems: Phase IV



VoiceXML

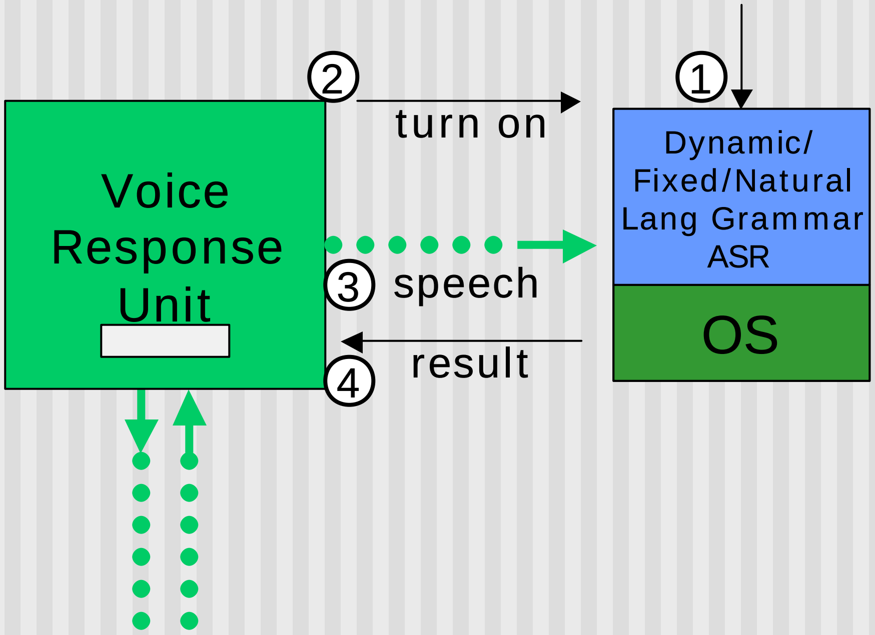
- At Version 2.0 of specification
- XML based markup language
- As HTML is to a web browser, VoiceXML is to a “voice browser”
  - With differences
- HTML is spatial while VoiceXML is temporal
  - Top-down execution paradigm with control flow
- VoiceXML is a programming language in itself
  - Javascript may be embedded in document
  - Complex applications can be created in a single document
- Mechanisms to invoke Javascript
- Fairly complex language
  - Experienced programmers can run into non-obvious problems

Speech Processing (ASR/TTS)

Overview of ASR

- Speech feed can be raw PCM or "feature-extracted" data samples

1. Load grammar
2. Turn on recognizer
3. Feed speech
4. Get result

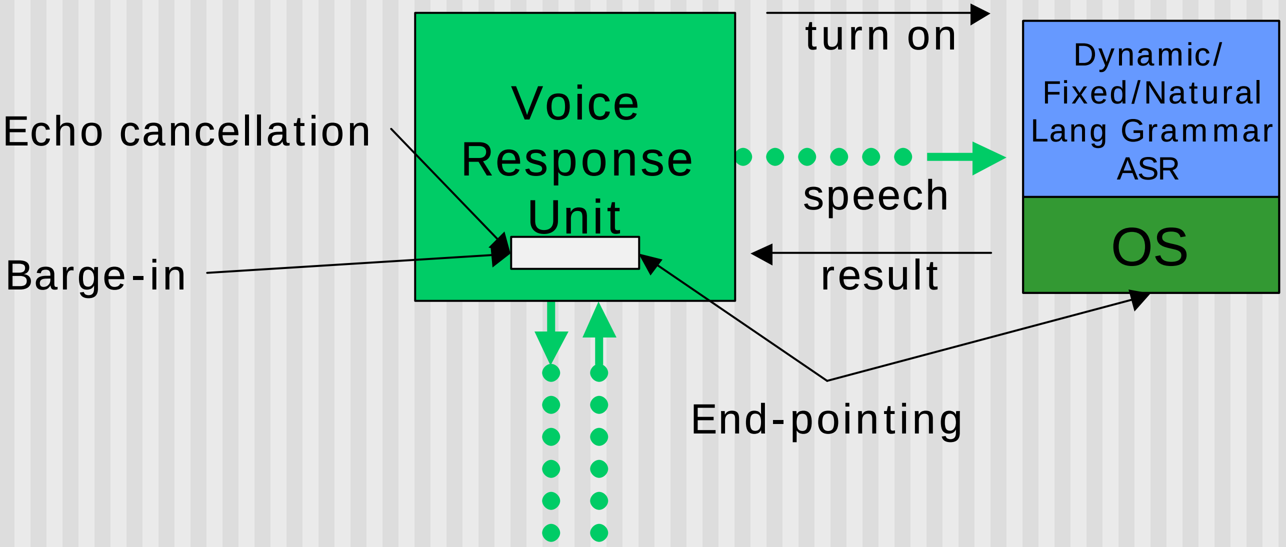


Attributes of Integration

- Echo cancellation
    - Subtraction of waveform of prompt being played
- End-pointing
    - Determination of when speech input is complete
- Barge-in
    - Immediate termination of prompt play when speech input is detected

Integration of ASR

- Echo cancellation
  - Use DSP resources to compute in real-time
- End-pointing
  - Can be in VRU or ASR engine
- Barge-in
  - Can be done in software or hardware

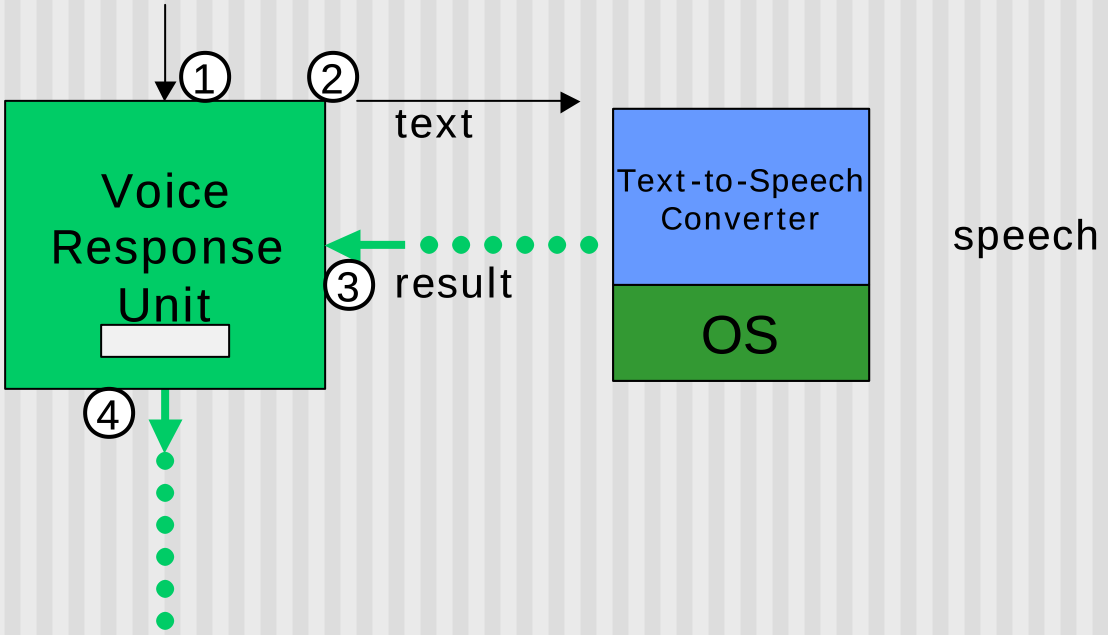


ASR Integration Mechanisms

- ASR SDK provided by Speech Recognition technology vendors
- Client/Server architecture commonly used
- IVR system developer uses the SDK to incorporate support for ASR
- ASR vendor implements licensing of speech recognition within the SDK supplied client
- Some ASR vendors do resource management of recognition server farms
- Until recently ASR vendors provided their own implementation on top of Dialogic/NMS/... boards
- With Voice over IP, this has become unnecessary

Overview of TTS

1. Text from Application
2. Send Text to Converter
3. Get resulting audio
4. Play audio



TTS Integration Mechanism

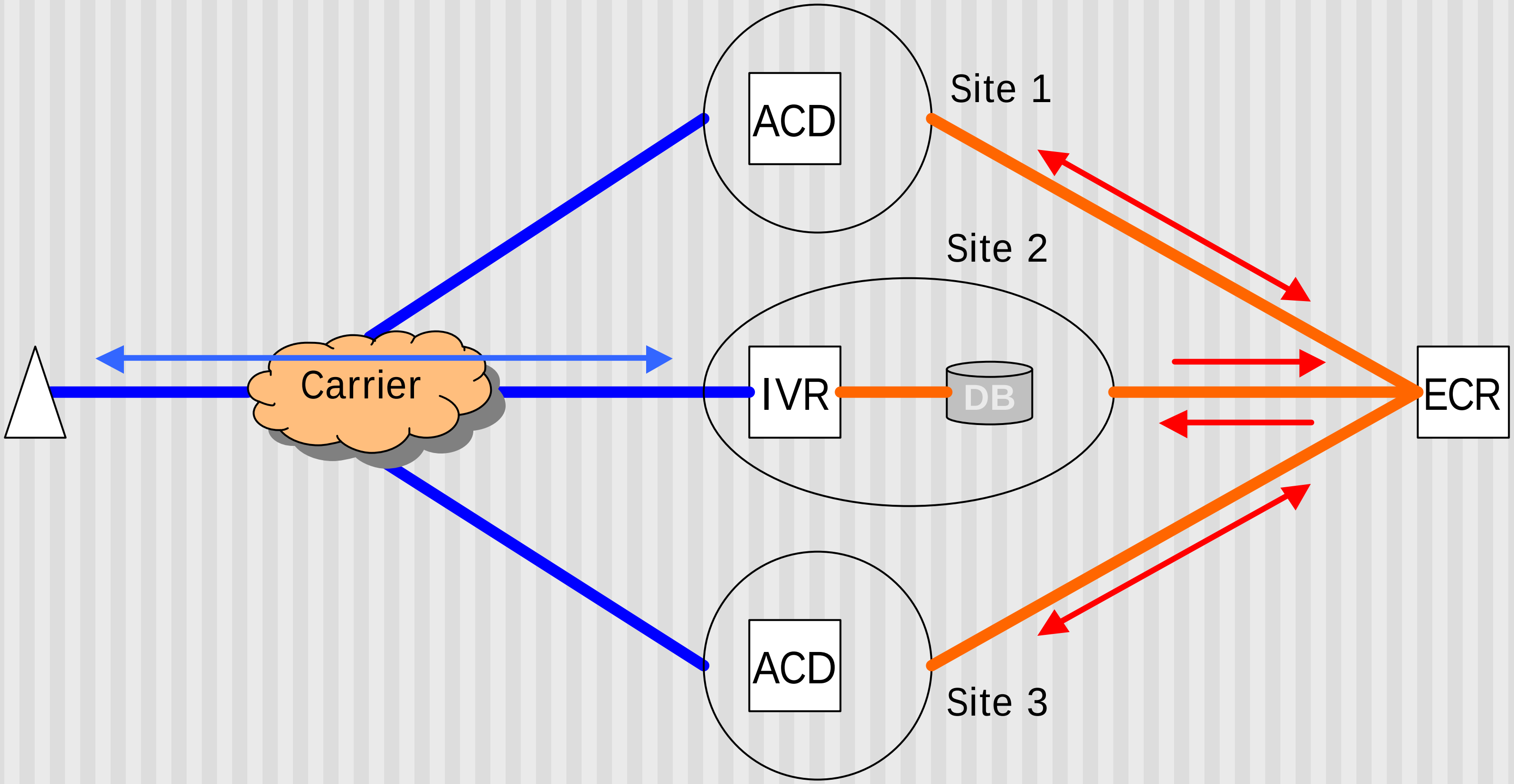
- Much simpler than ASR
- Generally through an on-board API provided by the TTS vendor
- Can be implemented on-board or as a server resource
  - Use a web service to do the conversion from text to audio

Computer Telephony Integration

Anatomy of a Call (Contact)

Automated Service	Queue	Talk
IVR System	ACD	Human
Learn more about the caller and their need	Queue to who can best serve	Provide customer service

Multi-site Contact Center

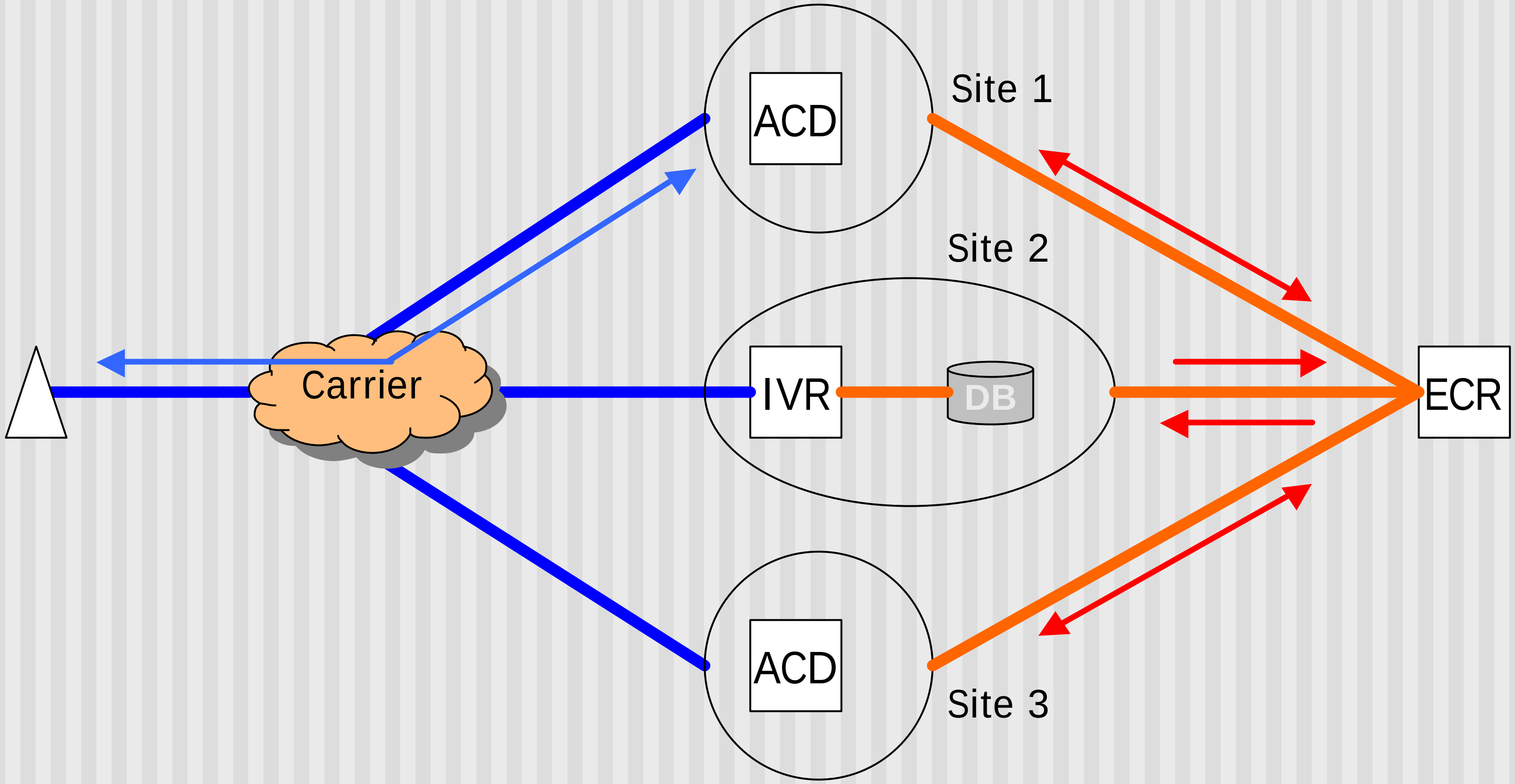


IVR = Interactive Voice Response System

ACD = Automatic Call Distributor

ECR = Enterprise Call Router (includes CTI Server)

Multi-site Contact Center

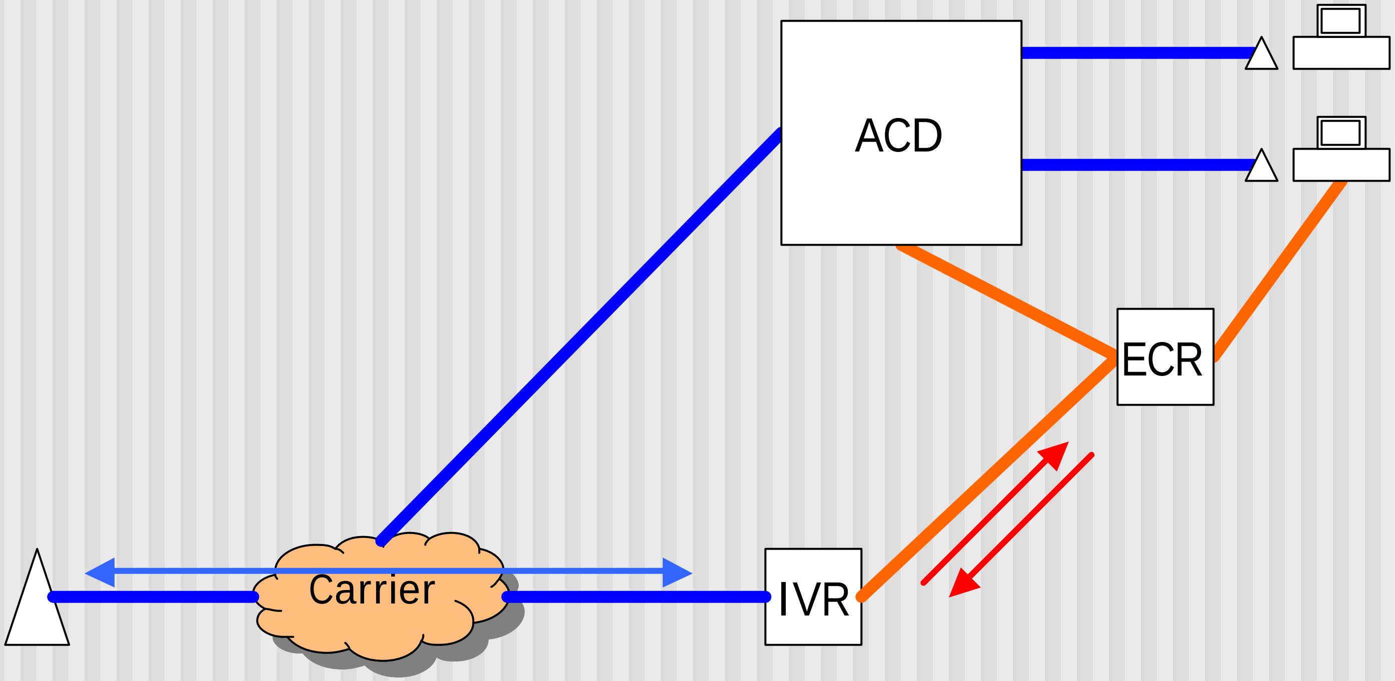


IVR = Interactive Voice Response System

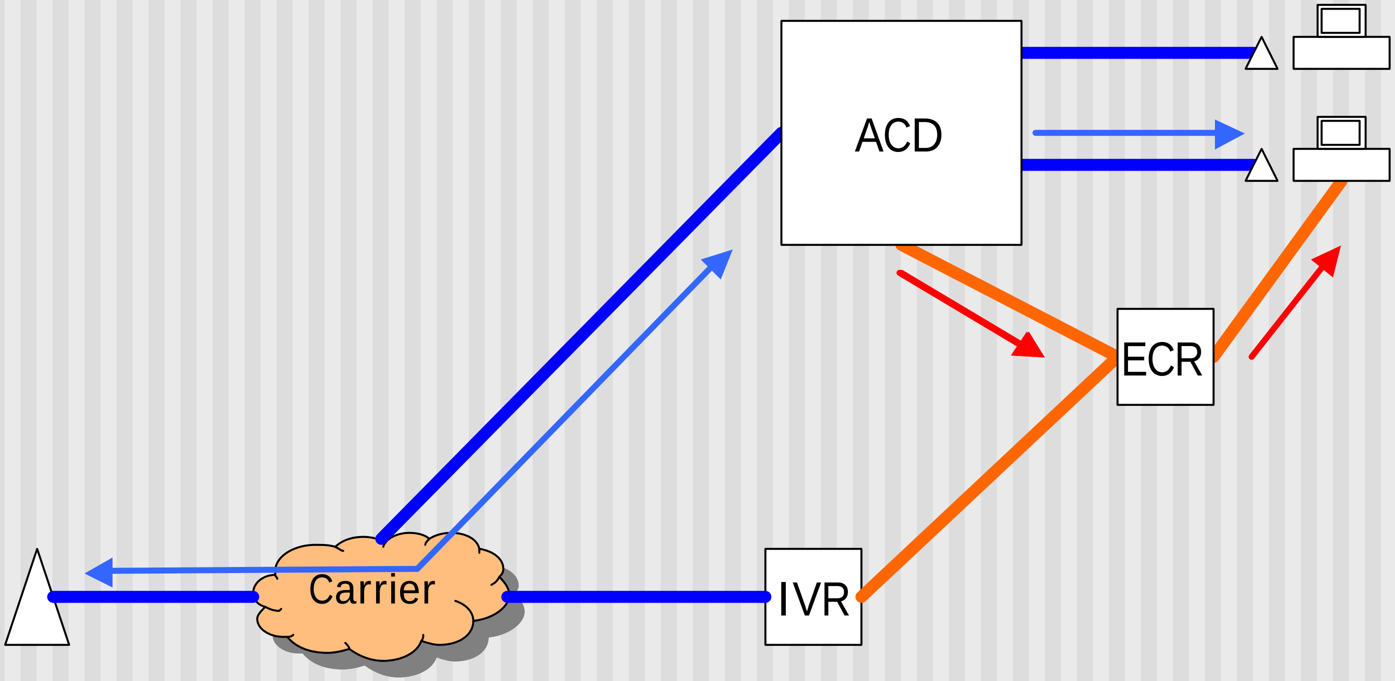
ACD = Automatic Call Distributor

ECR = Enterprise Call Router (includes CTI Server)

Inside a Contact Center Site



Inside a Contact Center Site



Call Router & CTI Server Role

- Enterprise Call Router manages across sites
  - Directs the call to the right location
  - Directs the call to the right skill
- ACD manages one switching fabric
  - Defines skill groups
  - Queues the call to the right group
  - Switches the call to the available rep
- CTI server relates call with associated data
  - Compiles the data, tracks call through its phases
  - Provides call data to agent desktop when call arrives
- Desktop performs screen pop
  - Uses data provided by CTI server to locate caller’s record in the CRM application

Challenges in TDM World

- Identifying the call that has been moved from point to point
- Carrier can only provide limited info with call
  - Automatic Number Identification (ANI)
  - Dialed Number Identification System (DNIS)
- Transfer of call associated data through specialized carrier interfaces
  - AT&T Transfer Connect with User to User Information (UUI) on ISDN trunks
- In practice, all the call associated data must be held in some repository
  - Role of of ECR & CTI software

IP Telephony

Signaling Standards

- Started with TDM signaling applied to IP
- H.323
  - Q.931 from TDM world transported over TCP/IP
  - Originally did not have a way to indicate DTMF
    - Caused major problems for DTMF based applications when using compression codecs
    - Added DTMF transport to H.323 v2
  - Protocol more telecom oriented than Internet protocols such as HTTP
- SIP
  - Started as an Internet oriented protocol (circa 96-97)
  - Has now become the de-facto signaling standard
  - Multi-media support (compared to H.323)

Media Standards

- RTP & RTCP
- Started as only media transport
- Various codecs
  - G.711, G.729a, G.723, etc.
- Evolved to incorporate support for in-band DTMF through typed RTP packets
  - RFC 2833

The Big Deal About SIP

- Simple, text-oriented syntax
- Request/response protocol similar to HTTP
- Easily decoded and understood
- Can transport other payload
  - Some Media Gateway vendors indicate DTMF via SIP messages
- No limits on payload size
- Transports call associated data during call setup
  - Support for ANI/DNIS
- Virtually eliminates the CTI problem at call setup
- Still, detractors point to number of messages during call setup

Standards Activities

VoiceXML 2.0

- A W3C standard
  - VoiceXML 2.0 entered Candidate Release status 1/28/03
- HTTP based markup language of communication between application and the voice platform
- Functional operations like prompt play, input, etc.
- Control flow and conditionals
- JavaScript for fine grain control by applications
- Definition of speech grammar language called SRGS
  - End of vendor dependent grammar definition languages
- Standardization of speech recognition operations

Standardization in VoIP

- Philosophy of loosely coupled application servers
- Treating speech engines as recognition servers
  - Clients send speech data as RTP packets
  - Results are returned based on recognition against currently active grammars
- SIP's extensibility
  - Definition of new operations
  - Carry variable payload
- SIP and VoiceXML are key standards in voice processing in the VoIP arena
  - Standard definition of how a VoiceXML start URI should be enclosed in the call setup payload

MRCP

- Media Resource Control Protocol (MRCP) draft created by Cisco, Nuance and Speechworks
- Cisco wanted to embed VoiceXML interpreter in media gateway & support speech recognition
- Nuance and Speechworks are leading speech recognition vendors in North America
- IETF Draft created in late 2001
- Products becoming available in 2003 from several speech vendors

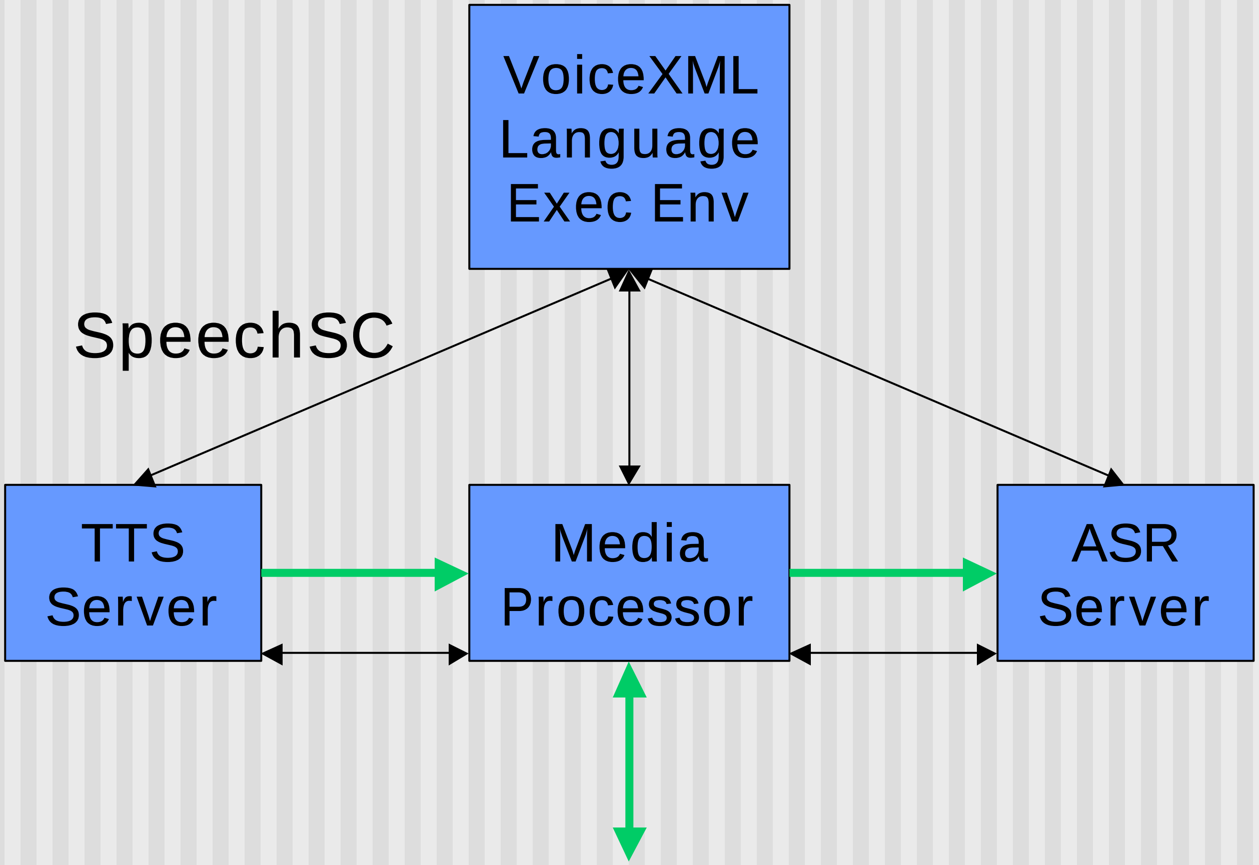
MRCP Elements

- Methods (Controller to Recognition Engine)
  - SET-PARAMS
  - GET-PARAMS
  - DEFINE-GRAMMAR
  - RECOGNIZE
  - GET-RESULT
  - RECOGNITION-START-TIMERS
  - STOP
- Events (Recognition Engine to Controller)
  - START-OF-SPEECH
  - RECOGNITION-COMPLETE

SpeechSC

- SpeechSC: Speech Services Control
- Working Group chartered by the IESG in mid 2002
  - Create an inclusive forum for MRCP type of standard
- Broader scope than MRCP
- Includes Speaker Identification and Verification
- Should come out with a draft protocol during 2003

SpeechSC Model



VoiceXML Language Exec Env

SpeechSC

TTS Server

Media Processor

ASR Server

Standards Reference

- VoiceXML
  - http://www.w3.org/TR/voicexml20/
- SIP
  - RFC 3261
  - http://www1.ietf.org/ids.by.wg/sip.html
  - http://www1.ietf.org/ids.by.wg/sipping.html
  - http://www.sipcenter.com/
- MRCP
  - http://www1.ietf.org/internet-drafts/draft-shanmugham-mrcp-03.txt
- SpeechSC
  - http://www1.ietf.org/ids.by.wg/speechsc.html

Impact on Industry and Customers

IVR & Speech Vendors

- VoiceXML changes the IVR landscape
- Proprietary hardware/software stacks under assault
  - Standard hardware building blocks now available for TDM call termination
  - Faster PCs increases density
  - Layering of software taking place
- IVR vendors are responding with language interpreters
  - But forgetting the web-architecture
- VoiceXML commoditizes speech technology
  - End to proprietary grammar languages
  - Focus on application design (PS)
  - Good enough speech recognition
  - Some speech vendors now competing with partners

Enterprises

- Enterprises asking for VoiceXML in RFPs
- See the value of competition in choosing voice delivery platform
  - Application creation no longer requires the IVR vendor to provide the development tools
  - Can now leverage web infrastructure for voice
- But, still want the graphical drag and drop tools offered by IVR vendors
  - 3rd parties tools are slowly coming to market
- Complexity is now in the integration of the IVR platform with their contact center infrastructure

Service Providers

- Service Providers already there
- IP infrastructure and VoiceXML/SIP relationship drives new voice infrastructure plans
- Separation of applications from delivery platform enables them to provide value added managed services
- Until now Service Providers could offer only limited hosted voice processing
  - Menus in the cloud still a mega business
- VoiceXML allows Service Providers to offer “fat-minutes” during which the Enterprises customer has dynamic control over caller interaction
  - Also address the SMB market
- MRCP/SpeechSC allows deployment of multiple ASR engines

Outlook

Trends

- Application vendors
  - Sell “packaged” applications
  - Until VoiceXML came along, this was not a feasible business
  - Climate brutal towards such vendors
- Development tools vendors
  - Application server for voice deployment
  - Also a tough business
- System Integrators
  - Reduction of costs through reuse of web trained resources
  - Build in-house components for reuse in client projects

Continuing Challenges

- Building a good voice user interface is still the real challenge
- Penetration of speech recognition technology is still around 10-15%
- Deployment of IP technology by Carriers is not a given
  - Though enhanced voice services is a catalyst
- Enterprises are still spending on customer service
  - But tight with budget

Agenda

- Interactive Voice Response
- Speech Processing
- Computer Telephony Integration
- IP Telephony
- Standards Activities
- Impact on Industry and Customers
- Outlook