AI and the Future of Customer Engagements
Customer service today

Web-based support is experiencing a cognitive bot revolution
✓ Making it cheaper and faster to handle common queries
✓ Handling queries 24 hours a day, 365 days a year

Call center challenges
✓ Voice is still the primary contact method used by customers
✓ Cost of optimization (live calls cost from $6 - $10 per call)
✓ Wait time reduction
✓ Improve customer satisfaction
✓ Easier access to internal documents and info

Increase
When a customer call is resolved the customer cross-selling acceptance rate is increased by 20%

Improve
A 1% improvement in FCR = 1% improvement in customer satisfaction
Call Center Transformation through Artificial Intelligence

- AI can handle calls through automation previously handled only by humans
- AI can reduce customer call hold times with live agents through agent assist
- AI can convert large volumes of unstructured data into consumable content
- AI can surface insights and trends occurring across a contact center

By 2018 50% of agent interactions will be influenced by real-time analytics

By 2020 the use of Virtual Agents will increase to handling 10% of business to consumer first level engagements

Example of a customer engagement with a Cognitive Contact Center

1. Customer calls support
2. Call routed to Voice Agent (BOT).
3. Voice and SMS exchanged with Caller.
4. Voice Agent publishes call details to context store.
5. Call transferred to live agent.
6. Media from live call forked to Agent Assistant.
7. Agent pulls call details from context store.
8. Agent Assistant publishes insights to live agent in real-time.

CTI Context Store

Enterprise Telephony System

Agent Assistant Voice Agent

Cognitive Contact Center

©2018 IBM Corporation
What is a **Voice Agent**?

- **Voice Agents** (or Voice BOTs) answer common questions over the phone using various AI services to...
  - **Reduce** calls going to human agents
  - **Resolve** queries more quickly
  - **Shorten** queue times for human agents

- **Voice Agents** are a combination of:
  - A set of **AI services** that:
    - **Understand** natural language
    - **Extract** caller temperament and emotions
    - **Process** unstructured data for consumption
  - A **Cognitive IVR** that:
    - **Connects** to VoIP network
    - **Orchestrates** AI services and other APIs
    - **Publishes** AI infused reports to track KPIs
AI services used by Voice BOTs

- **Speech To Text** – transcribe audio into text
- **Text To Speech** – synthesize text into audio
- **Natural Language Classifier** – intent mapping
- **Tone Analysis** - analytical, confident, professional, cheerful, polite, formal, or friendly
- **Sentiment Analysis** - positive, negative or neutral
- **Deep Search** – ingest and search unstructured data for long tail responses
What is a Cognitive IVR?

- Orchestrates various AI services to simulate a live agent
- Integration point for CTI, call control, reporting and business logic
- Interfaces to telephony network (SIP, RTP)
Cognitive IVR (Voice and AI Orchestration)

Voice Sources
- SBC
- Asterisk
- Enterprise Telephony
- SMS Providers
- SIP Trunking Providers

Logging, Monitoring, Reporting
- ELK
- NoSQL
- Splunk
- Prometheus
- REST Server
  Static WAVs

AI Services
- Natural Language Understanding
- Tone Analysis
- Sentiment Analysis
- Deep Search
- Analytics

Speech Services
- Google Speech
- Watson Speech
- MRCP (Nuance)
- Amazon Poly

REST, etc.

SIP, RTP

Public and Private Cloud Fabrics (e.g. Kubernetes)

MRCP, WSS
## What’s the difference between a Cognitive IVR and Traditional IVR?

<table>
<thead>
<tr>
<th>Traditional IVR</th>
<th>Cognitive IVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily DTMF driven</td>
<td>Primarily Voice driven</td>
</tr>
<tr>
<td>Directed Dialog</td>
<td>Open ended dialog</td>
</tr>
<tr>
<td>Grammar driven</td>
<td>Language and Acoustic models</td>
</tr>
<tr>
<td>VoiceXML</td>
<td>Natural Language Understanding (NLU)</td>
</tr>
<tr>
<td>MRCP based speech servers</td>
<td>Cloud speech services (plus MRCP)</td>
</tr>
<tr>
<td>Web Service orchestration</td>
<td>AI service orchestration (e.g. NLU, Tone, Sentiment)</td>
</tr>
<tr>
<td>Traditional form factors such as OVAs</td>
<td>Cloud Native, Dockerized workloads</td>
</tr>
<tr>
<td>Traditional KPIs (e.g. containment, opt outs)</td>
<td>KPIs based on AI driven analytics (tone, topic modeling)</td>
</tr>
</tbody>
</table>
Voice Agent **Challenges** (plus resolutions)

### Speech to Text
- Understanding Dialects (acoustic model training)
- Domain specific language (language model training)
- Single syllable words and letters can be difficult to transcribe (grammars help)
- Noise filtering (media processing, Radisys MRF, etc.)
- CPU intensive (auto scaling, hardware assist)
- Voice Authentication (voice biometrics)

### Text to Speech
- To NOT sound robotic (SSML and Google Wavenet)
- CPU intensive (caching for static responses)

### Natural Language Understanding
- Collecting input from caller (Watson Assistant Slots)
- Intent training (pre-trained intents for industry verticals)
What is an Agent Assistant?

- **Agent Assistants** rely on real-time audio streams to help live agents utilize AI to:
  - Quickly answer customer questions
  - Reduce the time it takes an agent to complete a call
  - Improve customer satisfaction

- **Agent Assistants** are a combination of:
  - A set of **AI services** that:
    - Transcribe natural language
    - Map caller questions to intents and topics
    - Extract agent temperament and emotions
    - Search unstructured data to rapidly surface insights
  - A **Voice Gateway** that:
    - Taps into live phone calls
    - Transcodes media for consumption by AI services
    - Publishes text transcriptions to orchestration
AI services used for **Agent Assistant**

- **Speech To Text** – transcribe audio into text
- **Natural Language Classifier** – intent mapping
- **Tone Analysis** - analytical, confident, professional, cheerful, persuasive, polite, formal, or friendly
- **Sentiment Analysis** - positive, negative or neutral
- **Topic Modeling** – insights from unstructured data
What is an agent assistant Voice Gateway?

- A telephony device that bridges audio from phone calls to a set of orchestrated AI services
- Techniques include SIPREC, packet capture, conference call listener, Back-to-Back User Agents
Cognitive Agent Assistant Network Diagram

- **SIP Trunking Provider**
- **SBC (SIPREC client)**
- **Enterprise Telephony**
  - ACDs
  - SIP Session Management
- **SIPREC + RTP**
- **Voice Gateway (SIPREC Server)**
- **Speech Engine**
- **Service Orchestration**
- **REST Interface (text utterances)**
- **AI Services**
  - Analytics
  - Deep Insights
  - Tone Analysis
  - NLU
  - Intent Mapping
  - Sentiment
- **Custom APIs** (Systems of Record, etc.)
- **Context**
- **Publish insights to agent desktop**

**Cognitive Contact Center**
Agent Assist **Challenges** (plus resolutions)

- **Telephony Integration**
  - Tapping into phone calls (SIPREC, SIP Back-to-Back User Agent, MCU)
  - Delivering insights in real-time to an agent (Messaging via Kafka or MQTT)
  - Scaling to large numbers of concurrent calls (B2BUA, additional SBC ports)

- **Speech To Text**
  - Accurately transcribing speech into text (acoustic model training)
  - Speaker separation when using an MCU (speaker diarization)
Contact centers of the future will be cloud native

- **What is Cloud Native?**
  - Microservices
  - Containers from development to production
  - Continuous Delivery/Continuous Integration
  - Consistent fabrics across both public and private clouds
  - Multicloud - portability between clouds (IBM Public Cloud, AWS, etc.)

- **Why Cloud Native?**
  - Faster code delivery and tighter control over what is being delivered.
  - Massive yet dynamic scalability
  - End-to-end deployment pipelining
  - Portability
Containers, Docker and Kubernetes enables Cloud Native workloads

- Characteristics of typical Docker workloads in Kubernetes
  - Stateless
  - Immutable/ephemeral containers
  - Dynamically scalable

- What are the challenges with running VoIP applications in Kubernetes?
  - Real-time VoIP protocols like SIP and RTP are highly stateful
  - VoIP applications tend to be tightly coupled to the hardware
  - VoIP protocols not natively supported in Kubernetes
    - VoIP applications currently must run in hostNetwork mode
    - Requires external SIP load balancer of Kube PODs

- Where is all this heading?
  - Native support for SIP and RTP in Kubernetes
  - Kubernetes ingress router that supports SIP and RTP needed
Demo

- You can dial this number to talk to a Voice BOT about IBM Voice Gateway:
  - 1-855-969-4241

- Demo was built using IBM Voice Agent with Watson:
  - Uses Watson Speech To Text, Watson Text To Speech and Watson Assistant
  - [https://console.bluemix.net/catalog/services/voice-agent-with-watson](https://console.bluemix.net/catalog/services/voice-agent-with-watson)
Thanks!

IBM