Call Center Multimodal Voice Search

Matt Yuschik, Ph.D.
Human Factors Specialist,
Human Computer Interactions
Relationship Technology Management

Session A103 – Voice User Interface in Voice Search
Voice Search Conference
March 2-4, 2009
Convergys Corporation
A Global Leader in Relationship Management

Worldwide Capabilities
- 75,000 employees
- 85 contact centers, and clients in 70+ countries

A Leading Public Company
- $2.8 billion in revenues
- Listed on NYSE, S&P 500, Fortune 1000
- A Fortune Most Admired Company for eight consecutive years

Key Facts About Convergys
- Host more than 1 billion customer interactions annually
- Support more than 3 million employees and retirees worldwide
- Billing for 350+ million communications subscribers worldwide
- Top 10 for Innovative Use of Technology of InformationWeek (2007)
Relationship Technology - Call Center Solutions

**Customer Needs**  
(Per Forrester)  
Consistent Channel Experience  
Advanced Self-Service Capabilities  
Live e-channels Support  
Agents Who Know My History  
Proactive Notification and Support  
Extended Hours of Service

**Multi-Channel Self-Care**

Outside-In Approach to match customer needs to client needs

**Client Needs**  
(Per Yankee Group)  
Enhanced Customer Experience  
Improved Revenue Opportunities  
Lower Operational Costs

**Automated Self-Care**  
Virtual/Secret Agents (Self-Care Optimizer)

**Real-time Decisioning**  
Dynamic Decisioning Solution

**Web Solution**  
Speech Solution  
Mobile Device Solution
Path from Multimodal Agent to End-User Self-care

Point Solutions vs. Solution Path

- Transactions
- Dialogs
- Tasks
- Subtasks

- Multimodal presentation
- Conversational Interface
- Speaker Verification for security
- Hidden Agent for difficulties
- Device-Independent Multichannel

70 k SMEs
Solutioners
Observers
MM testers
Multimodal Preferences

What devices and combination of modes are available? What tasks can be done with multimodality, and which should?

<table>
<thead>
<tr>
<th>device</th>
<th>landline</th>
<th>cellphone</th>
<th>PC</th>
<th>PDA</th>
<th>3G phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input speak</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Input type</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Input tap</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Input GPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Output listen</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Output listen - TTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Output read text</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Output view figures</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Output view video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

**Input:**

What should be spoken?
What should be typed or pressed?

**Output:**

What should be read?
What should be heard?
Needs Assessment and Solution- Delivery Service

Outside-In Approach = Monitor Agents and Callers

Start with GUI–based Service Handling

Many layers / screens to complete a call

Monitor Services -> Transaction types / Use Cases

Identify most frequent functions

Decompose into Tasks

Navigation, Form Filling, Data Caching, Search

Solution = A Multimodal Agent Tool that includes Voice Search

Caller value-add

Navigation, Data-driven search
Better Customer Care Experience

Agent value-add

Improves Agent Productivity and Quality
Better Agent Satisfaction, increased retention
Simulation Results - Longitudinal

Longitudinal testing
- Gives agents time to learn
- Identifies time training

Test Conditions –
- 4 Agents (6 mo -10 yrs experience)
- 1 UI per day for 3 days
- 5 Services
- 7 repetitions / svc – expect leveling
- No caller involved

UI Types – “wrapper” concept
- Existing GUI
  - Narrow – voice activate existing flow (AA)
  - Broad – streamline, follow agent/caller dialog (A2)
Pilot Test – Learning Types

Excellent Learners
“get it” quickly
<7 days

Normal Learners
accomodate then decrease
10-15 days

Delayed Learners
At risk
Don’t want to “get it”
Multimodal AHT - Performance Groups

AVA Group - w/o CTI Agents

Change from Baseline

Days on AVA

Faster (n = 7) | Little Change (n = 6) | Slower (n = 5)

Successful Agent

Comfortable with new technology
Shorter tenure
Uses most of the MMUI most of the time
Uses conference mode

Mismatched Agent

Longer tenure
Resistant to technology changes
Extensive use of Mute mode
Dependent on a notepad or scratchpad
Multimodal UI Design Principles – Voice Search

- Speech and graphics are both active at the same time
  - GUI and VUI are integrated, and do not just work in parallel
  - Error conditions are handled in both modalities
- Shortcuts for common transactions (mixed initiative dialog)
  - Multi-word utterances – (data driven phrases)
  - Timeout when default step is next – time to accept/reject
  - Limited acknowledgement
- The agent and the system are focused on one vocabulary, the same vocabulary
  - Don’t need SLMs or NL for the most part (constrained domain)
- Provide a temporary memory for data and speech
Conclusions and Next Steps

Wrapper requirements and effects
API approach

Component Applications
Voice Pad

Evolution to Multi-Modal Self-care devices

Application Assessment Process
Determining best places to add another modality (voice)

Categories of Transactions
Customer service hot-spots
Categorizing flows and tasks

Training
What should be learned, and when
Useful tools and practices
Questions, Comments, Insights?

Matt Yuschik, Ph.D.
Human Factors Specialist,
Human Computer Interactions
Relationship Technology Management
Multi-Channel Self-Care Solutions
Convergys Corporation
matt.yuschik@convergys.com