Tools and Toolkits for Voice and Animated Character-based Interventions

Timothy Bickmore & Daniel Schulman
College of Computer & Information Science
Northeastern University

Overview

• Tim
  – Dialogue-based Interventions
  – IVR languages & tools
  – ECA tools
• Dan
  – Dtask & LiteBody
Dialogue-based Behavioral Informatics

Current status of IVR systems

• Speech Recognition: accuracy of 85-90%+
• Speech Synthesis: high quality for short utterances
• Dialog Management: strongly-managed dialog flow
• Language Processing: Very limited
• Language Generation: Very trivial
Current IVR Apps

- Directory assistance
- Taxi bookings
- Stock transactions
- Remote banking
- Travel reservations
- Auto-attendants
- Pizza ordering
- And...

State of the Art: Scripted Interactions

- Scripts written by teams of experts
- Represented as flow charts or Augmented transition networks
- Implemented as state machines / ATNs / VXmL
Voice XML - What is it?

- Standardized language for specifying voice dialogs
- Output:
  - Prerecorded audio and text-to-speech (TTS)
- Input:
  - Touch-tone keys and Automatic Speech Recognition (ASR)
- Extension of XML
- Designed to interact with web-based applications

History

- 1995
  - Phone Web project by AT&T Research
- 1999
  - Lucent and AT&T have incompatible dialects of Phone Markup Language
  - So, VoiceXML Forum created with AT&T, Lucent, Motorola, and IBM
  - Team develops VoiceXML 0.9, a first pass at standardization
- 2000
  - VoiceXML 1.0 was created and submitted to World-Wide Web Consortium (W3C)
- 2001
  - VoiceXML 2.0 by W3C's Voice Browser Working Group
Big Picture:
The phone as a web browser

Intro to VXML
VXML is a kind of XML

– Tags and body
  
  ```xml
  <cmu>
    <welcome>Welcome to CMU! </welcome>
  </cmu>
  ```
  
  ```xml
  <ecom>
    <welcome>Welcome to the E-commerce!</welcome>
  </ecom>
  ```

– zero or more attributes
  
  ```xml
  <welcome accent="texan">Welcome</welcome>
  <welcome accent="pittsburgh">Welcome</welcome>
  ```

– Tag with no body
  
  ```xml
  <breath/>
  ```

A simple example

```xml
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.1">
<form>
  <field name="goingtoBeach" type="boolean">
    <prompt>“Are you going to Daytona Beach this year?”
    </prompt>
    <filled>Ohh...
      <if cond="goingtoBeach">That's great!
        <goto next="goingDocument.vxml" />
      </if>
      <else />
        bummer maybe next year.
        <goto next="notgoingDocument.vxml" />
      </else>
    </filled>
  </field>
</form>
</vxml>
```
A simple example
ASR grammar & TTS prompt

<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.1">
<form>
  <field name="goingtoBeach" type="boolean">
    <prompt>“Are you going to Daytona Beach this year?”</prompt>
    <filled>Ohh...
      <if cond="goingtoBeach">That’s great!
        <goto next="goingDocument.vxml" />
      </if>
      <else />
        bummer maybe next year.
        <goto next="notgoingDocument.vxml" />
      </if>
    </filled>
  </field>
</form>
</vxml>

A simple example
Execution conditioned on ASR recognition

<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.1">
<form>
  <field name="goingtoBeach" type="boolean">
    <prompt>“Are you going to Daytona Beach this year?”</prompt>
    <filled>Ohh...
      <if cond="goingtoBeach">That’s great!
        <goto next="goingDocument.vxml" />
      </if>
      <else />
        bummer maybe next year.
        <goto next="notgoingDocument.vxml" />
      </if>
    </filled>
  </field>
</form>
</vxml>
A simple example
if/else with goto

<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.1">
<form>
  <field name="goingtoBeach" type="boolean">
    <prompt>
      "Are you going to Daytona Beach this year?"
    </prompt>
    <filled>
      Ohh...
      <if cond="goingtoBeach">That's great!
        <goto next="goingDocument.vxml" />
      </if>
      bummer maybe next year.
      <goto next="notgoingDocument.vxml" />
    </filled>
  </field>
</form>
</vxml>

Commercial IVR Tools
IBM WebSphere Voice Toolkit
Commercial IVR Tools
Audium Studio

Commercial IVR Tools
Voxeo Designer
Commercial IVR Tools & Services
VoiceXML-based Solutions

Directory at:
http://www.voicexml.org

Embodied Conversational Agents

Automated Interventions
(not computer-mediated communication)
Motivation:
Acceptance & Usability

- Natural, intuitive modality provides rapid acceptance
- Emulate human face-to-face conversation
- Focus on nonverbal communicative behavior
  - gaze, posture, gesture, etc.
Motivation: Trust Building

• Agents can emulate human relationship-building behavior to build and maintain trusting, working relationships.

Motivation: Clients with Low Health Literacy

• Face-to-face communication, in conjunction with written instructions, is best

• However, most professionals have extremely limited time to spend with a customers
Current / Past Projects

- **NIH/National Cancer Institute**
  - R01 Oncology clinical trial adherence
  - R01 Exercise & UV for Cancer Prevention (URI)
  - R21 Exercise Promotion for Older Bilingual Latino Adults (Stanford)

- **NIH/National Institute on Aging**
  - R01 Clinic-based Walking Promotion for Older Adults (BMC)

- **NIH/National Heart, Lung, and Blood Institute**
  - R01 Hospital Discharge Patient Education (BMC)

- **NIH/National Library of Medicine**
  - R21 Exercise & Diet promotion
  - R21 Exercise promotion on mobile devices

- **AHRQ**
  - Web-based post-discharge medication adherence/screening (BMC)
  - Preconception care for young African American Women (BMC)

- **NSF**
  - Exercise promotion for older adults
Embodied Conversational Agents

• Dialogue Engine options:
  – Keyword match rules
  – HTN/ATN (VoiceXML)
  – Planner
• Output: typically TTS & animation
• Input options:
  – Multiple choice
  – Typed text
  – Speech

Embodied Conversational Agents
Output Toolkits
Embodied Conversational Agents Standardization Efforts

- SAIBA Framework
  - Situation / Agent / Intention / Behavior / Animation
- BML / FML


BML Synchronization Points
“Beginning with Windows® 7, Microsoft Agent will not be included or supported in future versions of the Microsoft Windows operating system. We encourage Microsoft Agent application developers and redistributors to evaluate their activities in light of this decision.”
Oddcast SitePal

- Scalable, web-based, 2D (Flash-based) characters.
- Limited dialog / logic support.

SmartBody (BML support)
USC Institute for Creative Technologies
3D Modeling & Animation

- Poser (Smith Micro)
- Maya
- etc.

Dialogue Engines

- Eliza/ALICE (e.g. alice.pandorabots.com)
- Latent Semantic Analysis
- Virtual People Factory (Virtual Experience Research Group, U. Florida)
- VoiceXML
- CEA-2018/
  DISCO
  web.cs.wpi.edu/~rich/cetask/
Dtask & Litebody

• Developed under NLM grant to provide open source tools for web-delivered, ECA-based interventions in behavioral medicine
• Demonstrate use of standards
• Ontologies & task models for exercise and fruit & vegetable promotion

• LiteBody – web-based ECA framework
• Dtask – plan-based dialog engine