Toward a Desirable Voice User Interface

Benjamin Ao
Alpine Electronics Research of America
19145 Gramercy Place, Torrance, CA 90501
1. State of Affairs

Fatal Driver Distraction vs Low VUI Utility Rate

- Fatal highway driver distraction on the rise
- VUI utility rate of in-car devices stays low (30%)

<table>
<thead>
<tr>
<th>Year</th>
<th>Crashes</th>
<th>Drivers</th>
<th>Fatalities</th>
<th>Distraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Crashes</td>
</tr>
<tr>
<td>2004</td>
<td>38,444</td>
<td>58,395</td>
<td>42,836</td>
<td>4,409 (11%)</td>
</tr>
<tr>
<td>2005</td>
<td>39,252</td>
<td>59,220</td>
<td>43,510</td>
<td>4,117 (10%)</td>
</tr>
<tr>
<td>2006</td>
<td>38,648</td>
<td>57,846</td>
<td>42,708</td>
<td>5,323 (14%)</td>
</tr>
<tr>
<td>2007</td>
<td>37,435</td>
<td>56,019</td>
<td>41,259</td>
<td>5,398 (14%)</td>
</tr>
<tr>
<td>2008</td>
<td>34,017</td>
<td>50,186</td>
<td>37,261</td>
<td>5,331 (16%)</td>
</tr>
</tbody>
</table>

Source: NCSA, FARS 2004–2007 (Final), 2008 (ARF)
1. State of Affairs

Reasons for Low VUI Utility Rate

- Manufacturer’s need for reliability resulting in supplementary use of VUI
- GUI-centric design leading to rigid hierarchical command structure
- Cognitive burden to memorize specific commands for various situations
- Low success rate and long effort time to accomplish user’s goal
- Natural human reluctance to talk with a mechanical device

Current Problem:
- Customer doesn’t feel natural
- Customer feels lost
- Customer doesn’t know how to get help

Requirements:
- User shall never feel confused, stuck, or rushed
- User can always get help
- VUI main role is assistant, not tour guide, it is supplementary technology to manual
- Must be useful to our customers

Actual problem report and solution requirement from same customer – highlight by author
2. The VUI-Centric Design Principle

The Speech Chain and Speech Communication

- Transmitting verbal message to get something done
- Conversation exchange and the turn-taking mechanism
- Interpretation and disambiguation with linguistic and worldly knowledge
- Error-handling or content-verification at phonetic and semantic levels
2. The VUI-Centric Design Principle

Model VUI Design after Human Speech Communication
3. VUI Design Considerations

Dialog Design Issues

- Domains and domain specificity – user convenience vs reco accuracy
- Turn-taking and initiation – human-initiated vs machine-initiated
- Error handling – recognition accuracy vs out-of-vocabulary commands
- Content verification and flexible dialog design – the meal-order scenario

Scenario 1: Multiple turns and multiple exchanges

**Waiter:** What would you like to order?
**Guest:** I'd have the seafood spaghetti.
**Waiter:** Would you like to have a salad to go with it?
**Guest:** Oh sure.
**Waiter:** What kind of salad? We have garden or Caesar.
**Guest:** Garden is fine.
**Waiter:** What dressing? We have ranch, Italian, thousand island, …
**Guest:** Italian, please.

Scenario 2: 2 turns and 1 exchange

**Waiter:** What would you like to order?
**Guest:** Seafood spaghetti and a garden salad with Italian dressing.
3. VUI Design Considerations

Training and assistance

• Offer training and/or assistance as guiding questions, e.g. Do you want A or B vs You can say A or B.

  Oh my gosh! What am I supposed to say? What am I supposed to do?

  Do you want to enter an address, go to a place, play music or make a phone call?

  Enter an address.

• Avoid unwanted assistance with detection of novice vs pro users, i.e. offer no assistance when user has learned a particular feature.

• Avoid annoying user with persistent assistance. No one likes uninvited and incessant advice. Knowing when to quit talking is intelligence.
4. Selection of Speech Engines

Importance of speech engine selection
• VUI usability critically depends on speech engine performance.
• Different engines have different capabilities.

ASR engine selection criteria
• Accuracy, latency, vocabulary size, grammar complexity, noise-robustness, speaker-independence.
• Need to test various engines with identical system setup and test material.

TTS engine selection criteria
• Phonetic accuracy, absence of acoustic artifacts, pleasant voice.
• Need to set up review panel to conduct listening test.
4. Selection of Speech Engines

ASR engine evaluation

TTS engine evaluation
Thank you!