Immersion in Multimodal Gaming

Playing World of Warcraft with Voice Controls

Tony Ricciardi and Jae min John
In a Sentence...

The goal of our study was to determine how the use of a multimodal control scheme affects a player's immersion in a video game.
Summary

What We Did:
- We compared 4 different control schemes for Blizzard Entertainment's *World of Warcraft*, 2 multimodal and 2 unimodal.

Why We Did It:
- Games are going multimodal.
  - Nintendo Wii, Nintendo DS, Rock Band, etc.
- How will this affect the player's experience?
Related Work

• "Multi-sensory Game Interface Improves Player Satisfaction but not Performance" - Nesbitt & Hoskens, 2008
  o involved game output, not player input
  o new modalities added new features, rather than replacing them

• "Use of Eye Movements for Video Game Control" - Smith & Graham, 2006
  o unimodal, but a new modality
  o improved player immersion

• "Multimodal Multiplayer Tabletop Gaming" - Tse, Greenberg, Shen & Forlines, 2007
  o multimodal board games are more fun and engaging than desktop video games (for collaborative play)
Why Warcraft?

1. The UI can be easily modified using the WoW API.

2. The game has two very distinct main tasks: controlling your character and communicating with your teammates. It was easy to assign different modalities to these tasks.
The Variables

**Independent:**

Communication Modality
- keyboard chat
- voice chat

Command Modality
- hotkeys
- voice commands

**Dependent:**

Attentiveness
- measured with eye tracking

Survey Responses
- perceived enjoyment, difficulty, presence, etc.
Communication Modality

Stop typing and start talking to your online friends.

Rent a Ventrilo server today!

Joined Channel: [5. GuildRecruitment - City]
[2. Trade] [Pilebunker]: WTS Various Epic Gems! PST for Info and Quantities! Also Cutting Epic and Blue gems, PST
[2. Trade] [Phattousai]: Booty Bay, then Gadget
[2. Trade] [Taintedbeef]: wts [Pattern: Volcanic Leggings] [Plans: Volcanic Hammer] 5g for both
[2. Trade] [Pascal]: WTS [Formula: Enchant Weapon - Potency]
[Edgemaster's Handguards] pm me
Command Modality
The Prototype

• Voice Commander
  o maps voice to button presses
  o created by Lenja of wow.curse.com

• We wrote some macros and preloaded them into the hotkey bar using a Lua script.

• When using voice communication, teammates spoke to each other directly, separated by a screen.
Our Hypothesis

• Using a multimodal interface will reduce a player's cognitive load.
  o "...users self manage their working memory limits by distributing information across multiple modalities..." - Oviatt, 2006

• This will decrease the cognitive investment required for initial engagement in the game.
  o "To lower the barriers to enter this level, the gamer needs to invest time, effort, and attention." - Brown & Cairns, 2004

• Thus, the player will become more immersed in the game when using a multimodal interface than when using a unimodal interface.
Experimental Design

We used a within-subjects design.

**Pros**
- requires fewer subjects
- easier to compare data

**Cons**
- a subject's experience using one interface may be influenced by previous interfaces

<table>
<thead>
<tr>
<th></th>
<th>Hotkey Controls</th>
<th>Voice Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyboard Chat</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Voice Chat</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
Experimental Design (cont.)

1. Subject creates character.

2. 10 minutes per case.
   • Cases were selected in random order.
   • Face recorded with camera.
   • One of us played with the subject as a teammate.

3. Short survey after each case.
   • Helped remove immersion before the next case.

4. After all 4 cases, the subject reported his/her favorite.
Eye Tracking

- Fixation duration has been shown to correlate with immersion in video games. - Tijs, 2006

- We had 2 parameters: gaze tilt and gaze pan.

- From these we calculated a 3-D gaze vector.

- Gaze vector used to calculate fixation duration:
  - The number of frames during which the gaze vector did not change by more than 20 degrees.

- Mean fixation duration calculated for each case, normalized for each subject.
Data Analysis (Mean Fixation Duration)

<table>
<thead>
<tr>
<th>Type</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>v chat/v con</td>
<td>.4144</td>
<td>10</td>
<td>.41965</td>
</tr>
<tr>
<td>v chat/k con</td>
<td>.2846</td>
<td>10</td>
<td>.32901</td>
</tr>
<tr>
<td>k chat/v con</td>
<td>.8347</td>
<td>10</td>
<td>.27077</td>
</tr>
<tr>
<td>k chat/k con</td>
<td>.3504</td>
<td>10</td>
<td>.34639</td>
</tr>
<tr>
<td>Total</td>
<td>.4710</td>
<td>40</td>
<td>.39699</td>
</tr>
</tbody>
</table>

$p < 0.05$

<table>
<thead>
<tr>
<th>(I) type</th>
<th>(J) type</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>k chat/v con</td>
<td>v chat/v con</td>
<td>.4203'</td>
<td>.15454</td>
<td>.010</td>
<td>.1069 - .7338</td>
</tr>
<tr>
<td>v chat/k con</td>
<td>v chat/v con</td>
<td>.5501'</td>
<td>.15454</td>
<td>.001</td>
<td>.2366 - .8635</td>
</tr>
<tr>
<td>k chat/k con</td>
<td>v chat/v con</td>
<td>.4843'</td>
<td>.15454</td>
<td>.003</td>
<td>.1709 - .7977</td>
</tr>
</tbody>
</table>

LSD
### Multiple Comparisons

<table>
<thead>
<tr>
<th>Dependent Variable: value</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LSD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v chat / v con</td>
<td>.1297</td>
<td>.15454</td>
<td>.407</td>
<td>-.1837 (.4431)</td>
</tr>
<tr>
<td>k chat / v con</td>
<td>-.4203’</td>
<td>.15454</td>
<td>.010</td>
<td>-.7338 (-.1069)</td>
</tr>
<tr>
<td>k chat / k con</td>
<td>.0640</td>
<td>.15454</td>
<td>.681</td>
<td>-.2495 (.3774)</td>
</tr>
<tr>
<td>v chat / k con</td>
<td>-.1297</td>
<td>.15454</td>
<td>.407</td>
<td>-.4431 (.1837)</td>
</tr>
<tr>
<td>k chat / v con</td>
<td>-.5501’</td>
<td>.15454</td>
<td>.001</td>
<td>-.8635 (-.2366)</td>
</tr>
<tr>
<td>k chat / k con</td>
<td>-.0658</td>
<td>.15454</td>
<td>.673</td>
<td>-.3792 (.2477)</td>
</tr>
<tr>
<td>k chat / v con</td>
<td>-.0640</td>
<td>.15454</td>
<td>.681</td>
<td>-.3774 (.2495)</td>
</tr>
<tr>
<td>v chat / k con</td>
<td>.0658</td>
<td>.15454</td>
<td>.673</td>
<td>-.2477 (.3792)</td>
</tr>
<tr>
<td>k chat / v con</td>
<td>-.4843’</td>
<td>.15454</td>
<td>.003</td>
<td>-.7977 (-.1709)</td>
</tr>
</tbody>
</table>
Data Analysis (Survey)

$p > 0.2$
Data Analysis (Preference)

- Voice Chat: 45%
- Voice Chat Control: 27%
- Keyboard Chat: 18%
- Keyboard Chat Control: 9%
Conclusion (Eye Tracking Data)

• We were right about voice control + key chat but wrong about key control + voice chat. Why?

• Maybe voice control + key chat lowered the cognitive barrier just enough, but key control + voice chat did too much.
  o Voice chat is MUCH easier than key chat in WoW.

• Cognitive underload has been shown to decrease attention significantly.
  o Young & Stanton, 2002

• The number of modalities is not enough: we must consider the type of modalities as well.
Conclusion (cont.)

Too Easy  Key Chat + Voice Controls  Key Chat + Voice Controls  Voice Chat + Key Controls  Too Hard

Unimodal
Conclusion (Survey Data & Preferences)

• Not statistically significant, but subjects generally preferred voice chat.

• We believe this is the result of problems with our design.
  o We were collocated, so voice chat was more natural.
  o We knew the subjects - they knew our voices.

• Still, the results weren't even close. This suggests that subjective and objective measures of immersion cannot be used interchangeably.
Future Work

- Improving our design
  - between-subjects
  - teammates who don't know each other
  - teammates not collocated

- Figuring out the right amount of cognitive load to achieve immersion.
  - The number of modalities.
  - The assignment of modalities to tasks.

- Investigating the relationship between subjective and objective measures of immersion.
Fin!