Speech-Driven Animation Constrained by Appropriate Discourse Functions

Najmeh Sadoughi, Yang Liu, and Carlos Busso

MOTIVATION

Background:
- Rule-based:
  - defining rules for behaviors based on the contextual information
  - repetitive behaviors
  - desynchronization between gestures and speech
- Speech-driven:
  - use of prosodic features to model behaviors
  - modeling emphasis, emotion, and timing of behaviors
- may not properly respond to the underlying discourse functions in the dialog

Proposed Solution:
- Create a bridge to fill the gap between speech-driven and rule-based systems

RESULTS

Statistical Analysis (MEAN)

<table>
<thead>
<tr>
<th>Discourse Function</th>
<th>Pitch</th>
<th>Roll</th>
<th>Pitch Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question vs. Non-Question</td>
<td>(F(1,452)=8.58) (p=0.004)</td>
<td>(F(1,452)=7.05) (p=0.008)</td>
<td>(F(1,452)=7.05) (p=0.008)</td>
</tr>
<tr>
<td>Affirmation vs. Non-Affirmation</td>
<td>(F(1,464)=7.87) (p=0.005)</td>
<td>(F(1,464)=10.42) (p=0.001)</td>
<td>(F(1,464)=6.74) (p=0.0097)</td>
</tr>
<tr>
<td>Negation vs. Non-Negation</td>
<td>(F(1,419)=5.17) (p=0.023)</td>
<td>(F(1,419)=4.99) (p=0.026)</td>
<td>(F(1,470)=4.30) (p=0.038)</td>
</tr>
<tr>
<td>Statement vs. Non-Statement</td>
<td>(F(1,470)=4.30) (p=0.038)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Question” vs. “Affirmation”

- 56% preferred C-jDBN3 over jDBN3
- 95.5% probability that this proportion is greater than chance
- Similar results for other questions

“Affirmation” vs. “Negation”

- 57% preferred jDBN3 over C-jDBN3
- C-jDBN3 closer to original videos
- Similar results for other questions

DISCUSSION

Conclusions:
- The statistical analysis demonstrated significant changes in behaviors across different discourse functions
- For “Question” we see more preference for C-jDBN3, while for “Affirmation” the results are not conclusive
- Perception of head motion dominate the evaluation
- “Affirmation” constraint is less effective since affects eyebrow

Future Work:
- We need more data to further explore this research direction
- Better talking heads

References:

This work was funded by NSF (IIS 1352950)