Integrating Speech Recognition into AAC Technology

Susan Koch Fager, PhD, CCC-SLP
Tom Jakobs, PE
David Beukelman, PhD, CCC-SLP

Background/Rationale

• Desire to use natural speech is innate
  – Automatic
  – Source of identity
  – Allow for more natural timing in interaction
  – Able to “hold the floor” compared to device-mediated interactions

• AAC technology tends to serve as a “replacement” for speech

Prototype description

• Speech recognition based on models of dysarthric speech
  – SSR (Supplemented Speech Recognition)
  – Incorporates speech, first letters of spoken words are typed, word prediction


• Forward facing monitor
  – User essentially has subtitles, can turn on or off

• Synthesized speech output if desired/needed
How the prototype works

- User types the first letter of the target word
- They speak the word
- The SSR attempts to recognize the word
  - If recognized, it is inserted in the line of text
  - If not, the word may appear in the word prediction list and the user can select if from there
  - Or user can spell the word out letter-by-letter
- What is written is displayed on the forward-facing monitor to the listener

Supplemented Speech Recognition

1. Automatic speech recognition based on models of dysarthric speech
   - System is further customized by individual user
2. First letter identification (alphabet supplementation)
3. Word prediction

How SSR functions

- User types the first letter of the target word
- User says the word
- Word shows up in line of text (most probable)
  OR
- Word is available on one of 6 word prediction buttons (next 6 probable word options) OR
- User has to type the word
SSR video


Evaluation

- The goal of the evaluation is to assess how this new method of supporting an AAC interaction impacts the listeners' behavior.
- Hypotheses:
  - Listeners' engagement as measured by on-task behavior and eye-gaze will increase during the RealTalk condition compared to traditional AAC condition

Conditions

- Traditional AAC (no speech, just text to speech with word prediction)
- RealTalk (prototype AAC system that incorporates supplemented speech recognition)

Participant

- Speaker with dysarthria
  - Female with CP, 74% sentence intelligibility, research assistant
- 5 listeners
  - 1 male, 4 females
  - 2 students in speech pathology, 1 accounting professional at rehab hospital, 1 IT manager at rehab hospital, 1 administrative assistant at rehab hospital
Tasks

- Unstructured conversation (introduction, hobbies, pets/vacations)
- Structured barrier tasks
- Tasks randomized per condition, per listener

Analysis

- Qualitative analysis of gaze behavior of listener:
  - All interactions video-recorded, transcribed, timed and coded for gaze behavior (on topic vs. off topic)
  - On topic- focused on speaker or task
  - Off topic- looking around room, focusing gaze on other objects in environment, engaging in conversation with 3rd party
- Proportion of words per participant (unstructured conversation)
- Qualitative feedback on comfort and preference of technology being used in a communicative interaction with the speaker

Results - Listener on-topic vs. off-topic

Results - Proportion of words used in unstructured conversation
Preference and qualitative comments

- All listeners preferred RealTalk to the traditional AAC condition for day-to-day communication with a communication partner
- Comments from listeners of Participant 1:
  - When she was talking it seemed more helpful during the conversation.
  - I didn’t know what to do or look at during the [traditional AAC] part. I felt uncomfortable.
  - I was able to focus and pay attention when she was talking.
  - I felt like I knew more what was going on when I could hear her talk.
  - The time delay in the traditional AAC condition felt unnatural. It felt more natural to be able to listen to her speech during the interaction.

- Comments from listeners of Participant 2:
  - Felt uncomfortable in Traditional AAC condition
  - Didn’t know how to react (where to look or attend while waiting)
  - Liked being able to look at monitor during RealTalk to stay focused on conversation
  - Felt as though the speaker had to work harder in the AAC only condition
  - Felt as though the AAC only condition took more time

3rd Party Listener of Interactions

- Asked typical individuals to view video-taped interactions of an individual with complex communication needs using traditional AAC and then using RealTalk prototype
- Mixed quantitative/qualitative design
  - Viewers completing ratings
  - Discussed why they rated the interactions the way they did

Preliminary data from 3 listeners

<table>
<thead>
<tr>
<th></th>
<th>Traditional AAC</th>
<th>RealTalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>How enjoyable did the conversation appear to be for the AAC user?</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>How enjoyable did the conversation appear to be for the non-AAC user?</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>How well did the AAC user seem able to say what she wanted to say?</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>How well did she seem able to say things in good time?</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>How well did she seem able to have control of the conversation?</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>How confident did she seem?</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>How relaxed did she seem?</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>How likely did she seem?</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>How well did the conversation flow?</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>How “natural” did the conversation seem?</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Overall, how satisfactory was the conversation?</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
Qualitative themes/supporting comments

- All 3 listeners chose RealTalk over Traditional AAC condition
  - Interactions appeared more comfortable for the listener
  - Less wait time
  - Listeners appeared more engaged

Theme 1: Listener comfort

- All 3rd party listeners felt that the listener (non-AAC user) in the interactions appeared to be more comfortable during the RealTalk condition compared to the Traditional AAC condition:
  - “She seemed less tense..more natural listening to the person using AAC”
  - “He seemed more at ease and focused on the conversation.”
  - “In the (Traditional AAC condition) she didn’t seem to know where to look….”

Theme 2: Flow of conversation

- All 3rd party listeners felt that the “flow” of the conversation was better during the RealTalk condition compared to the Traditional AAC condition:
  - “Both were slow…but the (RealTalk condition) appeared to flow more naturally.”
  - “There appeared to be more back and forth during the interaction.”

Theme 3: Enjoyment of interaction by the AAC user

- Many of the 3rd party listeners commented that the AAC user appeared to enjoy the interaction more when using RealTalk compared to the Traditional AAC condition:
  - “She was more animated...laughed more.”
  - “She seemed like she was more interested in the interaction...”
Acknowledgement

• Funding:
  – National Institutes of Health, NIDCD, Effective Self Expression for People with Severe Speech Disorders, 1R43DC012734-01
  – H133E140026 from the NIDLRR, U.S. Department of Health-RERC on AAC