Evaluating a method for automatic and objective scoring of verbal response for the Montreal Cognitive Assessment (MoCA)

Liam D. Kaufman, MSc ¹, Aparna Balagopalan, MSc ¹, Jekaterina Novikova, PhD ¹, and Fariya Mostafa, MSc ¹
(1) Winterlight Labs, Toronto, ON, Canada

Methods

- Scoring a multi-component verbal cognitive assessment from an audio recording requires differentiating the administrator from the participant (diarization), generating a transcript (automatic speech recognition; ASR), identifying the boundaries of each subtask in the transcript (task splitting), using the transcript of participant verbal responses to score the subtask (automatic scoring), and combining the subtasks into a total score.

- For this study, we chose to focus on evaluating automatic task segmentation approaches and the downstream scoring accuracy of the delayed recall subtask of the MoCA.

- MoCA recordings from 50 individuals were taken from a longitudinal natural history study of older adults (aged 55-90), recruited from the community and independent living facilities in Canada and the US.

- Recordings were manually diarized, transcribed, segmented and scored to produce a gold standard reference dataset.

- To segment tasks, ASR-produced transcripts from the rater were matched to the standard administration script for the MoCA. The following alignment algorithms were evaluated:
  - Phonetic Alignment
  - Keyword matching
  - Keywords matching with timestamp information

- We tested a variety of transcription and task splitting algorithm combinations to determine the upper and lower bound of performance.

- Delayed recall score was calculated (max = 5) based on the ASR of participant’s responses.

- Task segmentation performance was measured by examining the proportion of correctly identified delayed recall task boundaries.

- Mean absolute error (MAE) in delayed recall score for a given strategy was also tested.

Results

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Transcription for participant</th>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Delayed Recall MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Manual</td>
<td>ASR</td>
<td>84.0%</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>ASR</td>
<td>Phonemic</td>
<td>84.0%</td>
<td>1.87</td>
</tr>
<tr>
<td>3</td>
<td>ASR</td>
<td>Keyword</td>
<td>78.0%</td>
<td>2.20</td>
</tr>
<tr>
<td>4</td>
<td>ASR</td>
<td>Keyword + timestamps</td>
<td>82.0%</td>
<td>2.05</td>
</tr>
</tbody>
</table>

- Using manual segmentation and transcripts, automated scoring of the delayed recall task (checking reported words against the word list) was 100% accurate.

- Phonemic alignment was the most accurate task splitting algorithm (84.0%).

- The addition of automated task segmentation alone imparted a mean error of 0.76/5 on the delayed recall task (Scenario 1).

- Adding ASR for the participant responses (i.e., words recalled) increased the error to 1.87/5 (Scenario 2).

- This suggests that errors in ASR impart more error into the final delayed recall score than automatic segmentation of task boundaries.

- Additional combinations using Keyword based algorithms (Scenario 3 & 4), did not surpass phonemic alignment.

Conclusions

The results of this proof of concept study show that transcribed audio recordings can be used to automatically calculate Delayed Recall scores on the MoCA. Using an ASR-based algorithm to automatically segment MoCA tasks resulted in a mean error of 1.87/5 pts in Delayed Recall scores. Together these results suggest that automatic segmentation and scoring of audio recordings of cognitive assessments is feasible and further work using larger datasets is needed to fine tune the algorithms and improve scoring accuracy.

References


