On the robust automatic computation of speaking and articulation duration in ALS patients versus healthy controls

Jackson Liscombe1, Alexander Drcampo3, Hardik Kothare1, Oliver Roele1, Michael Neumann1, Andrew Cornish1, Doug Habberstad1, David Pautler1, David Suendermann-Oeft1, and Vikram Ramanarayanan1,2

1Modality.AI, University of California, San Francisco
Contact: jackson.liscombe@modality.ai

Results
To optimize articulation boundary detection for calculation of prosodic features.

Objective
Automatic Detection Methods

Praat Sound: To TextGrid (silences)...
● min_pitch: 50–500Hz
● time_step: 0.0–5.0s
● silence_threshold: -(1–70)dB
● min_silent_interval: 0.1–1.0s
● min_sounding_interval: 0.1–1.0s

WebRTC VAD
● aggressiveness: 0–3
● frame_duration: 10–30ms
● padding_duration: 100–2000ms

Comparing Task-Tuned Predictors (AD)

Comparing Cohort-Tuned Predictors (AD and SD)

Comparing Overall-Tuned Praat Per Task (AD and SD)

The only difference in tuned parameters was silence threshold. For each, these default settings were best:
● min_pitch = 100Hz
● time_step = 0.0s
● Minimum silent interval (s): 0.1
● Minimum sounding interval (s): 0.1

For speaking duration, the default setting of -25dB was best but for articulation duration, a new setting of -36dB silence threshold was found to be optimal.

<table>
<thead>
<tr>
<th>Task</th>
<th>silence_threshold</th>
<th>min_silent_interval</th>
<th>min_sounding_interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>OQ</td>
<td>-29</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>A</td>
<td>-46</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>DDK</td>
<td>-37</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>SIT</td>
<td>-26</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>R</td>
<td>-35</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>S</td>
<td>-37</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

All speech and non-speech regions were marked using Praat. Speaking duration (SD) = time between start of first and end of last speech region. Articulation duration (AD) = sum of all speech regions. Inter-annotator agreement on 45 utterances from 4 sessions was 0.42s RMGE.

Automatic Detection Methods

Praat Sound: To TextGrid (silences)...
WebRTC VAD

Comparing Task-Tuned Predictors (AD)

Comparing Overall-Tuned Praat Per Task (AD and SD)

Comparing Cohort-Tuned Predictors (AD and SD)

Comparing Cohort-Tuned Predictors (AD and SD)

Comparing Task-Tuned Predictors (AD)

Control cohort: not diagnosed with ALS, ALSFRS-R score of 48, 951 turns from 86 sessions.

Bulbar cohort: diagnosed with ALS, ALSFRS-R Bulbar subscore < 12, 526 turns from 48 sessions.

For speaking duration, performance was better for bulbar users with a slightly lower silence threshold compared to control users (-24dB vs -26dB) and a minimum sounding interval of 0.05s instead of 0.1s for the control cohort.

For articulation duration, performance was better for bulbar users with a higher silence threshold compared to control users (-38dB vs -28dB) and a lower minimum silence interval of 0.1s compared to 0.4s for control.

Task Descriptions

OQ Have you had any challenges when speaking, salivating, or swallowing? If so, please briefly describe any difficulties.
A Please take a deep breath and then say “aaa” until you run out of breath.
DDK Please take a deep breath and say “pataka” over and over until you run out of breath.
SIT Please say, “The job provides many benefits.” [Repeated 5 more times with different sentences.]
R Please read the text aloud to me, to the best of your ability. [Participant shown text of passage about bamboo.]
S Please describe what you see happening in this picture. Please try to speak for at least one minute.

Corpus
2,231 turns from 195 sessions collected between July 28, 2020 and February 22, 2021 from users of NEMSI to assess symptoms of ALS.