Workload Assessment using Speech-Related Neck Surface Electromyography

H-WORKLOAD 2018

Aaron Novstrup
Stottler Henke Associates
anovstrup@stottlerhenke.com

Terrance Goan
Stottler Henke Associates
goan@stottlerhenke.com

Dr. James Heaton
Mass. General Voice Center
James.Heaton@mgh.harvard.edu
The Problem

- Robust real-time workload monitoring is difficult!

- Existing psychophysiological signals have limited utility:
  - Multiple measures are *sensitive* but *divergent* [Matthews et al., 2015]
  - Limited *specificity*
  - Limited *diagnosticity*
fnsEMG: A Solution?

Existing sEMG sensor designs

Notional sensor design
fnsEMG: A Solution?

- Face/neck surface electromyography (fnsEMG) can recognize and classify *emotional responses*.
  
  [Van Boxtel, 2010; Cheng & Liu, 2008; Tassinary & Cacioppo, 1992; Van Boxtel et al., 1983]
fnsEMG: A Solution?

- Face/neck surface electromyography (fnsEMG) can recognize and classify emotional responses. [Van Boxtel, 2010; Cheng & Liu, 2008; Tassinary & Cacioppo, 1992; Van Boxtel et al., 1983]

- It can also quantify neuromuscular activity related to the vocalization and articulation of speech. [Meltzner et al., 2018; Meltzner et al., 2017; Denby et al., 2010; Jou et al., 2006]
Intermuscular Beta Coherence

- **Coherence** is a frequency domain measure of the *linear dependency or strength of coupling* between two signals/processes.

\[
|R_{xy}(\lambda)|^2 = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda)f_{yy}(\lambda)}
\]

- **Intermuscular beta-band coherence** (15-35 Hz) is known to decrease during divided attention or reduced movement precision. [Kristeva-Feige et al., 2002]
Neck Intermuscular Beta Coherence

• **NIBcoh** can distinguish *healthy* vs. *strained* voice production.  
  [Stepp, Hillman, & Heaton, 2010 & 2011]

• It has also been shown to *decrease* during speech when *attention is diverted* to a primary non-speech task.  
  [Stepp, Hillman, & Heaton, 2011]
Hypotheses

• NIBcoh decreases under divided attention.
  
  o Confirm this finding of [Stepp et al., 2011], under a more conservative statistical model
  
  o Examine the impact of full-wave rectification on the EMG-EMG coherence analysis, responding to concerns raised by [Neto & Christou, 2010] regarding this common practice

• NIBcoh is correlated with task performance in a time-pressured mental arithmetic task with verbal responses.
NIBcoh Dataset

• Dataset from [Stepp et al., 2011]
• Neck surface electromyography (sEMG) was recorded over ventral neck strap muscles in 10 vocally healthy individuals during:
  o normal speech
  o static non-speech maneuvers
  o singing
  o “clear” speech (intentionally produced to maximize intelligibility)
  o hyperfunctional speech (mimicking a strained/stressed quality)
  o divided-attention speech (under heightened cognitive load by rapidly counting backwards from 100 by 7s)
A “divided-attention” task

100, 93, 86, 78, 71, 64, 57, 50, 42, 35, 28, ...
Data Analysis

- [Stepp et al., 2011] treated tasks as the unit of analysis (e.g., measuring coherence and sEMG RMS across entire recordings).
- We manually annotated the sEMG dataset for error commission (from speech recordings), and examined change in sEMG related to errors and response time.
  - Variables analyzed included NIBcoh, sEMG magnitude, and gradient, as well as acoustic features: average/peak amplitude, spectral roll-off, cepstral peak prominence, and sound intensity.

Example of raw and RMS sEMG signals during 20 sec of reading from a participant with high beta coherence (0.63).
Replication Study

- Compared EMG-EMG coherence across 6 conditions from [Stepp et al., 2011] with repeated measures ANOVA

- Unrectified vs. Full-wave Rectified (see [Neto & Christou, 2010])

- Beta band (15-35 Hz) vs. 100-150 Hz band
Replication Study: Results

<table>
<thead>
<tr>
<th>Coherence measure</th>
<th>Size of effect (generalized $\eta^2$)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-wave rectified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta band</td>
<td>0.120</td>
<td>0.0715</td>
</tr>
<tr>
<td>100–150 Hz</td>
<td>0.0892</td>
<td>0.0905</td>
</tr>
<tr>
<td>Unrectified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta band</td>
<td>0.176</td>
<td>0.0194‡</td>
</tr>
<tr>
<td>100–150 Hz</td>
<td>0.0651</td>
<td>0.146</td>
</tr>
</tbody>
</table>
Do Signals Covary with Errors?

• Unit of analysis: individual response (e.g., “86”)

• Standardized EMG/acoustic variables

• Compared mean values for correct vs. incorrect responses

  $H_0 : \mu_C = \mu_{-C}$
Do Signals Covary with Errors?

<table>
<thead>
<tr>
<th>Measure</th>
<th>Estimated difference</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIBcoh</td>
<td>0.505</td>
<td>2.81</td>
<td>&lt; 0.006</td>
</tr>
<tr>
<td>NIBcoh-rect</td>
<td>0.078</td>
<td>0.548</td>
<td>0.585</td>
</tr>
<tr>
<td>EMG magnitude (site 1)</td>
<td>0.432</td>
<td>-2.22</td>
<td>0.028</td>
</tr>
<tr>
<td>EMG magnitude (site 2)</td>
<td>0.532</td>
<td>-2.50</td>
<td>0.014</td>
</tr>
<tr>
<td>EMG gradient (site 1)</td>
<td>0.120</td>
<td>-0.772</td>
<td>0.44</td>
</tr>
<tr>
<td>EMG gradient (site 2)</td>
<td>0.064</td>
<td>-0.414</td>
<td>0.68</td>
</tr>
<tr>
<td>mean acoustic amplitude</td>
<td>0.852</td>
<td>-2.35</td>
<td>0.020</td>
</tr>
<tr>
<td>peak acoustic amplitude</td>
<td>2.52</td>
<td>-2.96</td>
<td>&lt; 0.004</td>
</tr>
<tr>
<td>sound intensity</td>
<td>0.345</td>
<td>-1.97</td>
<td>0.051</td>
</tr>
<tr>
<td>spectral roll-off</td>
<td>0.276</td>
<td>1.21</td>
<td>0.230</td>
</tr>
<tr>
<td>cepstral peak prominence</td>
<td>0.258</td>
<td>-1.67</td>
<td>0.097</td>
</tr>
</tbody>
</table>
Do Signals Covary with Response Time?

- Unit of analysis: individual response (e.g., “86”)
- Standardized EMG/acoustic variables
- Computed correlation with response time
  - $H_0 : \rho_{r,v} = 0$
<table>
<thead>
<tr>
<th>Measure</th>
<th>Pearson’s $r$</th>
<th>$t$-statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIBcoh</td>
<td>-0.196</td>
<td>-2.22</td>
<td>0.028</td>
</tr>
<tr>
<td>NIBcoh-rect</td>
<td>-0.078</td>
<td>-0.868</td>
<td>0.387</td>
</tr>
<tr>
<td>EMG magnitude (site 1)</td>
<td>-0.176</td>
<td>-1.98</td>
<td>0.050</td>
</tr>
<tr>
<td>EMG magnitude (site 2)</td>
<td>0.008</td>
<td>0.084</td>
<td>0.933</td>
</tr>
<tr>
<td>EMG gradient (site 1)</td>
<td>0.076</td>
<td>0.849</td>
<td>0.397</td>
</tr>
<tr>
<td>EMG gradient (site 2)</td>
<td>0.111</td>
<td>1.24</td>
<td>0.218</td>
</tr>
<tr>
<td>mean acoustic amplitude</td>
<td>-0.139</td>
<td>-1.56</td>
<td>0.122</td>
</tr>
<tr>
<td>peak acoustic amplitude</td>
<td>0.185</td>
<td>2.08</td>
<td>0.040</td>
</tr>
<tr>
<td>sound intensity</td>
<td>-0.223</td>
<td>-2.54</td>
<td>0.012</td>
</tr>
<tr>
<td>spectral roll-off</td>
<td>-0.005</td>
<td>-0.053</td>
<td>0.958</td>
</tr>
<tr>
<td>cepstral peak prominence</td>
<td>-0.219</td>
<td>-2.49</td>
<td>0.014</td>
</tr>
</tbody>
</table>
Caveats and Limitations

• Exploratory reanalysis of dataset *not* collected to assess sensitivity of NIBcoho to cognitive demands
  ○ Task demands were not experimentally manipulated
• Small sample size (n=10)
• Single task
• Sensitivity to outliers
• Unexplained correlations among variables
What Next?

- Collect EMG from multiple face/neck muscles
- Compare with conventional workload indicators
- Manipulate task complexity and “automation reliability” in Multi-Attribute Task Battery (MATB)
- Manipulate perceived risk
Discussion and Q&A

Thank You!
Phase I Dataset

- Intermuscular sEMG beta-band coherence (15-35 Hz) is known to decrease during divided attention or reduced movement precision (Kristeva-Feige et al., 2002)
  - sEMG coherence was significantly reduced in the heightened cognitive load speaking task and hyperfunctional speech

\[ |R_{xy}|^2 = \frac{|\text{cov}_{xy}(f)|^2}{\text{cov}_{xx}(f)\text{cov}_{yy}(f)} \]
Neuromuscular anatomical/physiological foundation

- Muscles controlling speech articulation and facial expressions are relatively superficial and therefore accessible for non-invasive sEMG
  - We have used a set of up to 11 surface recording locations, providing robust speech-related signals
  - sEMG activity is similarly strong whether speech is spoken normally or only “mouthed” (e.g. no voice)

Adapted from Selkirk, King & Drumright. Anatomy & Physiology for Speech, Language and Hearing 3rd Edition
Project Goal

Develop a **robust, reliable, and highly sensitive** model capable of **assessing the level of workload strain** and of **detecting/predicting overload** in real time in the context of human-automation interaction, enabling:

- Adaptive automation
- Cognitive countermeasures
- User interface affordances
Planned Innovations

• Demonstrate utility of face/neck sEMG for workload assessment
• Show improved specificity of sEMG relative to alternate measures
• Develop a minimally obtrusive set of face/neck sEMG sensors and sensing locations for the workload assessment application
• Identify a complement of unobtrusive sensors and measures that, together with sEMG, allows for robust and reliable assessment of workload strain.
Neuromuscular anatomical/physiological foundation

- Our prior work has shown that sEMG information is sufficient for accurate automatic speech recognition using as few as 4 sensor locations
  - Information redundancy allows limited channel failure
Neuromuscular anatomical/physiological foundation

We will add particular forehead and periocular sensor locations to capture eye blink and facial expressions

- **Orbicularis oculi (orbital portion)**
  - Eye closing muscles
    - Blink rate is correlated with stress/anxiety
- **Frontalis, corrugator supercilii, and procerus**
  - Surprise, scowling and frowning muscles
    - Active during negative emotional state
    - Active during surprise from unexpected outcomes
    - Amount of activity shown to relate to video game playing skill level (Weinreich, Strobach & Schubert, 2015)
Neuromuscular anatomical/physiological foundation

• Recording locations can be unilateral because head/neck muscle contraction are typically symmetrical (at least for speech)

• Some bilateral recording locations will allow us to measure intermuscular coherence
  o Coherence in the beta band (15-35 Hz) is known to decrease under stressed or divided-attention conditions (Stepp, Hillman & Heaton, 2011)
  o Beta coherence is believed to originate from the motor cortex, and is influenced by cognitive state
sEMG Sensor Technology

- Custom “mini” sensors were created by Delsys Inc. for our neck/face recording needs, which are now part of their product line
  - Designed to record from the relatively small muscles of articulation and facial expression
  - Light-weight and non-encumbering
  - Wireless communication to signal processing hardware
sEMG Sensor Application

- Sensor placement is relatively easy and typically lasts for hours
  - Clean, shaven skin is ideal
    - Beards preclude perioral placement
  - Skin exfoliation through tape “peeling” provides strongest signals
    - Clear desk tape works well
  - Sensors attach through disposable double-stick interface strips
  - No conductive gel or paste needed
sEMG Recording

- sEMG sensors communicate wirelessly 40 meters
- Multi-function design embeds Triaxial Accelerometry in standard-size sensors
- Up to 16 EMG channels and 48 accelerometer channels
- 8 hours of operation in full transmission mode (2 hour recharge time)
- 16-bit resolution, 1926 Hz EMG sampling rate
- USB Connection to PC (laptop or desktop)
- Real-time feedback of signal strength & battery status
Experimental Testbed / Tasks
Discussion

• Key challenges/opportunities from your perspective?

• Tasks / data sets

• Initial demonstration ideas and expectations

• Relationships to ongoing work

• Phase II proposal/award process