Using Articulation Index Band Correlations to Objectively Estimate Speech Intelligibility Consistent with the Modified Rhyme Test

IWCE 2014

Public Safety Communications Research

Steve Voran/Andrew Catellier
Electronics Engineer– ITS

Department of Commerce – Boulder Labs
HUH?
A Modified Rhyme Test (MRT) uses sets of rhyming words to measure intelligibility. Consider:

Please select the word _____.

- Went
- Sent
- Bent
- Dent
- Tent
- Rent
An MRT consists of asking lots of people lots of those questions using speech processed by a system under test.

For Example:
Please select the word _____.

- Went
- Sent
- Bent
- Dent
- Tent
- Rent
Please select the word _____.

- Went
- Sent
- Bent
- Dent
- Tent
- Rent
Please select the word _____.

- Hold
- Cold
- Told
- Fold
- Sold
- Gold
MRT Trial 2

Please select the word _____.

○ Hold
○ Cold
○ Told
○ Fold
○ Sold
● Gold
Please select the word _____.

- Pat
- Pad
- Pan
- Path
- Pack
- Pass
Please select the word _____.

- Pat
- Pad
- Pan
- Path
- Pack
- Pass
2008
http://go.usa.gov/ByTB

2010
http://go.usa.gov/ByTQ

2012
http://go.usa.gov/ByTw
Tested in 2008:

3 communication systems in 9 environmental conditions

(plus 3 conditions with another comm. system)

with speech from 6 talkers

50 lists of 6 sentences each

\[((3 \times 9) + 3) \times 6 \times (50 \times 6)\]

= 54,000 trials

1200 trials per listener = 45 listeners
MRT is very expensive
What if it were possible to estimate intelligibility using an automated system?
Automatic Speech Recognition (ASR) is a generic term for...speech recognition.

But the task at hand isn’t exactly speech recognition.
MRT by ASR  
(No Human Subjects)  
(Not your ordinary ASR)

http://go.usa.gov/Bmp9

Stephen Voran  
Institute for Telecommunication Sciences  
Boulder, Colorado, USA  
svoran@its.bldrdoc.gov  
www.its.bldrdoc.gov/audio
ASR Goals

- Large vocabulary
- Speaker independence
- Maximally robust to impairments
- Use Mel Cepstrum or Mel Cepstrum changes over time as features in hidden Markov models or deep neural nets

MRT Goals

- 6 words
- 4 (or 6) speakers
- Simulate human ability to comprehend
- Use Articulation Index Bands as features and correlate loudness patterns across time
ABC–MRT T–F Patterns
6 Original Words

Pat
Pad
Pan
Path
Pack
Pass
Match DUT Signal to One of Six Originals

DUT Signal

Pat

Pad

Pass

(and three others)
Time Alignment

DUT Signal

Pat
Time Alignment

DUT Signal

Pad
Time Alignment

DUT Signal

Pass
Time-Aligned T–F Patterns

DUT Signal

Pat
AI Band Correlations

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
</tr>
</tbody>
</table>
AI Band Correlations

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
</tr>
</tbody>
</table>
AI Band Correlations

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
<th>Pan</th>
<th>Path</th>
<th>Pack</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
<td>.33</td>
<td>.62</td>
<td>.58</td>
<td>.54</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
<td>.72</td>
<td>.43</td>
<td>.60</td>
<td>.59</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
<td>.59</td>
<td>.54</td>
<td>.46</td>
<td>.47</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
<td>.26</td>
<td>.11</td>
<td>.03</td>
<td>-.17</td>
</tr>
</tbody>
</table>
Select Highest Correlation in each AI Band

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
<th>Pan</th>
<th>Path</th>
<th>Pack</th>
<th>Pass</th>
<th>Winning Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
<td>.33</td>
<td>.62</td>
<td>.58</td>
<td>.54</td>
<td>Path</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
<td>.72</td>
<td>.43</td>
<td>.60</td>
<td>.59</td>
<td>Pan</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
<td>.59</td>
<td>.54</td>
<td>.46</td>
<td>.47</td>
<td>Pan</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
<td>.26</td>
<td>.11</td>
<td>.03</td>
<td>-.17</td>
<td>Pad</td>
</tr>
</tbody>
</table>
Correct Answer is “Pan”

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
<th>Pan</th>
<th>Path</th>
<th>Pack</th>
<th>Pass</th>
<th>Winning Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
<td>.33</td>
<td>.62</td>
<td>.58</td>
<td>.54</td>
<td>Path</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
<td>.72</td>
<td>.43</td>
<td>.60</td>
<td>.59</td>
<td>Pan</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
<td>.59</td>
<td>.54</td>
<td>.46</td>
<td>.47</td>
<td>Pan</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
<td>.26</td>
<td>.11</td>
<td>.03</td>
<td>-.17</td>
<td>Pad</td>
</tr>
</tbody>
</table>
## Find Success Rate

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
<th>Pan</th>
<th>Path</th>
<th>Pack</th>
<th>Pass</th>
<th>Winning Word</th>
<th>Correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
<td>.33</td>
<td>.62</td>
<td>.58</td>
<td>.54</td>
<td>Path</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
<td>.72</td>
<td>.43</td>
<td>.60</td>
<td>.59</td>
<td>Pan</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
<td>.59</td>
<td>.54</td>
<td>.46</td>
<td>.47</td>
<td>Pan</td>
<td>Y</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
<td>.26</td>
<td>.11</td>
<td>.03</td>
<td>-.17</td>
<td>Pad</td>
<td>N</td>
</tr>
</tbody>
</table>

\[
c = \frac{6}{17} = 0.35
\]
### Correct for Guessing

<table>
<thead>
<tr>
<th>AI Band</th>
<th>Pat</th>
<th>Pad</th>
<th>Pan</th>
<th>Path</th>
<th>Pack</th>
<th>Pass</th>
<th>Winning Word</th>
<th>Correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.55</td>
<td>.37</td>
<td>.33</td>
<td>.62</td>
<td>.58</td>
<td>.54</td>
<td>Path</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.55</td>
<td>.72</td>
<td>.43</td>
<td>.60</td>
<td>.59</td>
<td>Pan</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>.36</td>
<td>.41</td>
<td>.59</td>
<td>.54</td>
<td>.46</td>
<td>.47</td>
<td>Pan</td>
<td>Y</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>-.07</td>
<td>.30</td>
<td>.26</td>
<td>.11</td>
<td>.03</td>
<td>-.17</td>
<td>Pad</td>
<td>N</td>
</tr>
</tbody>
</table>

\[ c = \frac{6}{17} = 0.35 \]

\[ c' = \frac{6}{5} (c - \frac{1}{6}) = 0.22 \]
ABC–MRT Summary

• Perform time alignment
• Perform one correlation on each AI band for each of the 6 possibilities
• Take note of the word with the highest correlation in each AI band
• ABC–MRT result is the success rate across bands, averaged over lists, talkers
• How does ABC–MRT perform?
Development Database

- 28 conditions, \(0.02 \leq \text{MRT} \leq 0.84\)
- 4 types of speech coding, 7 acoustic environments

- Pearson Correlation 0.950
- RMSE 0.084
- No optimized parameters!
Development Database with Fitting

\[ \hat{\phi} = 0.865 \cdot c' + 0.119 \]

- Pearson Correlation 0.950
- RMSE 0.059
- 2 optimized parameters!
Add 3 Unseen Databases

- Total of 139 conditions, 71% of data unseen
- Unique acoustic conditions, speech coding, amplifier overload, radio channel impairments with analog and digital radios

![Graph showing correlation]

- Pearson Correlation 0.955
- RMSE 0.099
All Data with Fitting

\[ \hat{\phi} = 1.109 \cdot c' + 0.050 \]

- Pearson Correlation 0.955
- RMSE 0.073
- 2 optimized parameters!
## Compare with Other Options

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Correlation</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC–MRT</td>
<td>.96</td>
<td>.07</td>
</tr>
<tr>
<td>ABCa–MRT</td>
<td>.96</td>
<td>.07</td>
</tr>
<tr>
<td>Normalized Covariance Measure</td>
<td>.93</td>
<td>.09</td>
</tr>
<tr>
<td>Coherence Speech Intelligibility Index</td>
<td>.55 to .74</td>
<td>.17 to .21</td>
</tr>
<tr>
<td>PESQ</td>
<td>.84</td>
<td>.14</td>
</tr>
<tr>
<td>POLQA</td>
<td>.76</td>
<td>.16</td>
</tr>
</tbody>
</table>
Free Software

http://go.usa.gov/BvRQ

• The software runs relatively quickly (time will be dominated by data acquisition—it takes 34 minutes to pass all speech through a system under test)

• Can use fewer than 1200 files per condition

Graph shows deviation across 139 conditions

• Green is Max
• Blue is RMS
Work needed:

- Our results are narrowband (4 kHz) because all our test data is narrowband.
- In order to test ABC–MRT on wideband signals, an MRT using wideband signals must be created.
- Help us make this work better! Share ideas, data and results.
- Our MRT audio database and test results are available here: [http://go.usa.gov/BvRQ](http://go.usa.gov/BvRQ)
Caveats

- ABC–MRT provides intelligibility estimates
- There exist conditions for which ABC–MRT will produce less accurate results (for example, it is untested on wideband audio)
Questions
For Additional Information:
http://www.pscr.gov

Andrew Catellier
acatellier@its.blrdloc.gov