Vowel Space and Intelligibility in Dysarthric Speech

Heejin Kim
Beckman Institute, University of Illinois at Urbana-Champaign

Mark Hasegawa-Johnson
ECE, University of Illinois at Urbana-Champaign

Adrienne Perlman
SHS, University of Illinois at Urbana-Champaign

Supported by NIH R21-DC008090-A and NSF 05-34106
Dysarthric speech

Deviant speech patterns

- Imprecise consonants
- Hypernasality
- Excessive or monotonic intonation
- Vowel centralization
  - smaller vowel space
Dysarthric speech

- Vowel centralization

- F1 - tongue height, F2 - anterior/posterior tongue position
- Shown here are three corner vowels (F1/F2 extremes)
Dysarthric speech

- **Vowel centralization**
  - F1 - tongue height, F2 - anterior/posterior tongue position
  - Shown here are three corner vowels
Dysarthric speech

- Vowel centralization
  - F1 - tongue height, F2 - anterior/posterior tongue position
  - Shown here are three corner vowels

Liss et al., 2000; Turner et al., 1995; Weismer et al., 2001; Weismer et al., 2000; Tjaden et al., 2005
Previous work
Turner et al., 1995; Weismer et al., 2001; Tjaden & Wilding, 2004; McRae et al., 2002; Weismer et al., 2000

- Relationship btw the size of vowel space & intelligibility
  - Mostly wrt speaking rate modification
  - To measure perceived intelligibility, all used DME over speaking rate
  - Mostly corner vowels /i, a, u, æ/
  - ALS or PD group
  - Results: Wide range of measured effect
    - From 45% to 0%
In this study

- Relationship btw the size of vowel space & intelligibility
  - Mostly wrt speaking rate modification
    - speaker’s overall intelligibility
  - To measure perceived intelligibility, all used DME over speaking rate
    - word transcription
  - Mostly corner vowels /i, ɑ, u, æ/
    - both corner and noncorner vowels
  - ALS or PD group
    - CP
In this study

- We examined the F1 and F2 values of 6 vowels (/i, α, u/, /ɛ, ɛ, u/) produced by individuals with spastic cerebral palsy

- Research questions
  - Correlation between vowel space and intelligibility
    - Do speakers with lower intelligibility have smaller vowel space compared to those with higher intelligibility?
  - Is the effect of intelligibility different on corner vowel space vs. noncorner vowel space?
Database

- This study used a subset of our database, that has been collected for our project “Development of Automatic Speech Recognition for dysarthric speech”
  - Kim et al., 2008, Dysarthric speech database for universal access research, *Interspeech*

- Speakers
  - Recruited only subjects with cerebral palsy
    - Mostly spastic dysarthria
    - Subjects for this study were entirely spastic
Data collection: Equipments

- 8-microphone array
- Laptop computer
- Multi-channel digitizer
- Video camera
- Studio lighting at 750w

- Subjects were asked to read an isolated word displayed on a PowerPoint slide on a computer
- An experimenter sat next to the subject and advanced the PowerPoint slides
Data collection: Material

A total of 765

- 255 words in each of 3 blocks
  - 10 Digits (e.g. zero, one, two)
  - 26 Radio Alphabet Ws (e.g. alpha, bravo, charlie)
  - 19 Computer commands (e.g. delete, enter, tab)
  - 100 Common Words (e.g. the, and, go, no)
    - from the Brown corpus of written English
  - 100 Uncommon Words (e.g. naturalization, enthuse)
    - chosen from children's novels digitized by Project Gutenberg to maximize phone-sequence diversity

- 3 repetitions of 155 words
- 455 distinct words
Intelligibility rating

- **Purpose**
  - To acquire an index for severity of speech disorder

- **Material**
  - A total of 225 speech files / SPK
    - 200 distinct words from Block 2
    - 25 repeated words for intra-listener reliability assessment
  - Presented on a web page
Intelligibility rating

Listener
- 5 naive listeners / SPK
- Were asked to provide orthographic transcriptions of each word

Analysis
- The percentage of correct responses was calculated, and averaged across 5 listeners
- Intelligibility category
  - 0 % - 25%: very low
  - 26% - 50%: low
  - 51% - 75%: mid
  - 76% - 100%: high
Acoustic analysis

- Labeling
  - 6 Target vowels
    - 3 corner vowels /i, α, u/ (‘beat’, ‘Bob’, ‘boot’)
    - 3 noncorner vowels /ɪ, ɛ, u/ (‘bit’, ‘bet’, ‘book’)
  - Digits, radio alphabet and common words from all 3 blocks
  - Vowel mid point, using Praat
Analysis 1

- F1 and F2 (Bark)

- F1/F2 acoustic vowel space
  - To illustrate the degree of dispersion and overlap among vowels
  - An ellipse that covers 86.5%

- 1-way ANOVA
  - /i/ vs. /ɪ/; /ɪ/ vs. /ɛ/; /u/ vs. /ʊ/

- 6 speakers
  - 1 High, 1 Mid, 2 Low, and 2 Very Low
  - The number of vowel tokens of each speaker ranged from 223 to 227
    - The number differed across speakers due to vowel deletion and mispronunciation
Results: M09 (high, 86%)

<table>
<thead>
<tr>
<th></th>
<th>/i/ vs. /ɪ/</th>
<th>/ɪ/ vs. /ɛ/</th>
<th>/u/ vs. /ʊ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>F2</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>
Results: M05 (mid, 58%)

<table>
<thead>
<tr>
<th></th>
<th>/i/ vs. /ɪ/</th>
<th>/ɪ/ vs. /ɛ/</th>
<th>/u/ vs. /ʊ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>F2</td>
<td>**</td>
<td>**</td>
<td>NS</td>
</tr>
</tbody>
</table>
Results: M06 (low, 39%)

<table>
<thead>
<tr>
<th></th>
<th>/i/ vs. /ɪ/</th>
<th>/ɪ/ vs. /ɛ/</th>
<th>/ʊ/ vs. /u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>F2</td>
<td>**</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>
Results: M07 (low, 28%)

<table>
<thead>
<tr>
<th></th>
<th>/i/ vs. /I/</th>
<th>/I/ vs. /ɛ/</th>
<th>/u/ vs. /u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>F2</td>
<td>**</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

F1 (Bark)

F2 (Bark)
Results: F03 (very low, 6%)

<table>
<thead>
<tr>
<th></th>
<th>/i/ vs. /I/</th>
<th>/I/ vs. /ɛ/</th>
<th>/u/ vs. /ʊ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>NS</td>
<td>**</td>
<td>NS</td>
</tr>
<tr>
<td>F2</td>
<td>NS</td>
<td>**</td>
<td>NS</td>
</tr>
</tbody>
</table>
Results: M04 (very low, 2%)
Results summary

- Speakers with lower intelligibility had
  - a larger degree of overlap between vowels
  - a smaller vowel space
Analysis 2

- To compared the size of corner vowel space with that of noncorner vowel space as a function of intelligibility

- The area of triangular vowel space was defined using the mean F1 and F2 values of each vowel
  - The size was calculated using an equation
    - $\Delta = \frac{1}{2}(-x_2y_1+x_3y_1+x_1y_2-x_3y_2-x_1y_3+x_2y_3)$,
      - $x_1, x_2, x_3 =$ mean F1 values
      - $y_1, y_2, y_3 =$ mean F2 values
Results

![Bar chart showing vowel space area (Bark²) for different speakers.](image)

- **Vowel space area (Bark²)**
  - **Speaker**: M09, M05, M06, M07, F03, M04
  - **Legend**:
    - Red: corner
    - Blue: non-corner
Conclusion

- As intelligibility decreases,
  - vowel space reduced
  - overlapping among vowels increased

- The size of corner vowel space substantially decreased, as intelligibility decreased
  - the effect on noncorner vowel space was not as salient as on corner vowel space
Conclusion

Our findings suggest that expansion of corner vowel space would yield considerable gain in intelligibility in the remediation of dysarthric speech

Ongoing analysis

- Vowel formants for more speakers including control subjects
- Articulatory error patterns for consonants
Available dataset as of today

- **Recording**
  - 17 SPKs (plus 4 from our prelim study)
  - 13 control speakers

- **Audio**
  - A total of 5355 speech files / SPK
    - 255 words x 3 blocks x 7 microphones
  - Each wav file has one isolated word

- **Video**
  - 765 video clips (255 x 3 blocks) / SPK

- **Available via sftp upon request**
  - contact authors (hkim17@illinois.edu; jhasegaw@illinois.edu)
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Age</th>
<th>Speech Intelligibility (%)</th>
<th>Diagnosis</th>
<th>Motor Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>M01</td>
<td>18</td>
<td>Very low (15%)</td>
<td>spastic</td>
<td>Uses a wheelchair</td>
</tr>
<tr>
<td>M04</td>
<td>18</td>
<td>Very low (2%)</td>
<td>spastic</td>
<td>Uses a wheelchair, AAC &amp; head devices</td>
</tr>
<tr>
<td>M05</td>
<td>21</td>
<td>Mid (58%)</td>
<td>spastic</td>
<td>Uses a wheelchair</td>
</tr>
<tr>
<td>M06</td>
<td>18</td>
<td>Low (39%)</td>
<td>spastic</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
<tr>
<td>M07</td>
<td>58</td>
<td>Low (28%)</td>
<td>spastic</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
<tr>
<td>M08</td>
<td>28</td>
<td>High (93%)</td>
<td>spastic</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
<tr>
<td>M09</td>
<td>18</td>
<td>High (86%)</td>
<td>spastic</td>
<td>Ambulatory, able to sign</td>
</tr>
<tr>
<td>M10</td>
<td>21</td>
<td>High (95%)</td>
<td>mixed</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
<tr>
<td>M11</td>
<td>48</td>
<td>Mid (62%)</td>
<td>athetoid</td>
<td>Uses a wheelchair, used stamp to sign</td>
</tr>
<tr>
<td>M12</td>
<td>19</td>
<td>Very low (7%)</td>
<td>mixed</td>
<td>Uses a wheelchair, signed with help</td>
</tr>
<tr>
<td>M13</td>
<td>44</td>
<td>(Very low)</td>
<td>spastic</td>
<td>Uses a wheelchair, signed with help</td>
</tr>
<tr>
<td>M14</td>
<td>40</td>
<td>(high)</td>
<td>spastic</td>
<td>Ambulatory, able to sign</td>
</tr>
<tr>
<td>M16</td>
<td>24</td>
<td>Low (43%)</td>
<td>mixed</td>
<td>Uses a wheelchair, signed with help</td>
</tr>
<tr>
<td>F02</td>
<td>30</td>
<td>Low (29%)</td>
<td>spastic</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
<tr>
<td>F03</td>
<td>51</td>
<td>Very low (6%)</td>
<td>spastic</td>
<td>Uses a wheelchair and AAC</td>
</tr>
<tr>
<td>F04</td>
<td>18</td>
<td>Mid (62%)</td>
<td>athetoid</td>
<td>Uses a wheelchair</td>
</tr>
<tr>
<td>F05</td>
<td>22</td>
<td>High (95%)</td>
<td>spastic</td>
<td>Uses a wheelchair, able to sign</td>
</tr>
</tbody>
</table>
Thank you!