The effects of vocal characteristics on the recognition of concurrent syllables

Martin Vestergaard

Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG
Cocktail Party

- The “Cocktail party problem” (Cherry, 1953)
  - It’s not a problem for normally hearing listeners
- Factors that facilitate listening to speech in an environment with distracting speech
  - Level differences
  - Temporal asynchronies
  - Speaker differences vocal characteristic
Vocal Characteristics

- In natural communication sounds, there are three important kinds of information:
  - resonator shape
  - glottal pulse rate
  - resonance (rate)

- Body size
- Speaker identity
- Age
- Gender

Pulse rate
Resonance rate
Experiment design

• Task
  • Identify the syllable in the interval that stays lit
  • Indicate answer in the graphical user interface
Stimuli – vocal characteristics

- Target speech
  - **Pitch** 172 Hz
  - **VTL** 15 cm
- Distracter speech
  - **Pitch** (137 – 215 Hz)
  - **VTL** (11 – 21 cm)
Stimuli – voiced and whispered speech

Voiced /ti/

Whispered /ti/

Waveform

Spectrum

F1

F2

F3

F1

F2

F3
# Stimuli – constraints

(matching temporal envelope according to phonemic specification of syllables)

<table>
<thead>
<tr>
<th>M</th>
<th>N</th>
<th>L</th>
<th>R</th>
<th>W</th>
<th>Y</th>
<th>B</th>
<th>D</th>
<th>P</th>
<th>G</th>
<th>T</th>
<th>K</th>
<th>S</th>
<th>F</th>
<th>V</th>
<th>Z</th>
<th>SH</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ma</td>
<td>na</td>
<td>la</td>
<td>ra</td>
<td>ya</td>
<td>da</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>me</td>
<td>ne</td>
<td>le</td>
<td>re</td>
<td>ye</td>
<td>be</td>
<td>pe</td>
<td>ge</td>
<td>te</td>
<td>ke</td>
<td>se</td>
<td>fe</td>
<td>ve</td>
<td>ze</td>
<td>she</td>
<td>he</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>mi</td>
<td>ni</td>
<td>li</td>
<td>ri</td>
<td>yi</td>
<td>bi</td>
<td>pi</td>
<td>gi</td>
<td>ti</td>
<td>ki</td>
<td>si</td>
<td>fi</td>
<td>vi</td>
<td>zi</td>
<td>shi</td>
<td>hi</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>mo</td>
<td>no</td>
<td>lo</td>
<td>ro</td>
<td>yo</td>
<td>bo</td>
<td>po</td>
<td>go</td>
<td>to</td>
<td>ko</td>
<td>so</td>
<td>fo</td>
<td>vo</td>
<td>zo</td>
<td>sho</td>
<td>ho</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wu</td>
</tr>
</tbody>
</table>

- Target triplet: **da wu osh**
- Masker triplet: **ti na ez**
- Concurrently at 0 dB SNR
- Pre-cursor, 0 dB SNR
Stimuli – constraints

(matching temporal envelope according to phonemic specification of syllables)

Target triplet

da

wu

osh

Masker triplet

ti

na

ez
Results

- Effects of pitch and size
Results

- Effects of voicing

Vestergaard et al, 2007

Vestergaard et al, 2005
Results

- Effects of direction (ITD)

Ives et al
Results

• Effects of temporal asynchrony
Results summary

- Vocal specification
  - When voices are similar they are harder to segregate than when they are different
  - Listeners benefit from pitch as well as size difference
  - Trading relationship between VTL and GPR is 1.9
- Whispered speech
  - Whispered speech is more disturbing than voiced speech
  - At low SNR whispered speech is more robust than voiced speech
- Spatial cues
  - Listeners benefit from ITD when size difference is small
- Temporal asynchrony
  - Listeners tend to hook on to unmasked consonants
  - Listeners benefit from glimpses particularly when size difference is small
Finish

- Thanks to
  - Nick Fyson
  - James Tanner
  - Sami Abu-Wardeh
  - Andy Taylor
  - Tim Ives
  - Roy Patterson

- Research supported by
  - UK-MRC (G0500221, G9900369)
  - EOARD (FA8655-05-1-3043)