Burst Spectrum as a Cue to Stop Consonant Voicing

English Production and Perception Results

Eleanor Chodroff and Colin Wilson
Johns Hopkins University
voice onset time
F1 onset
F1 transition
F_{0} contour
relative amplitude of aspiration
following vowel duration

spectral shape of the burst:
*lower frequencies* for voiced stops

Cues to stop consonant voicing
“Since most of our lax [voiced] stops were pronounced with vocal-cord vibration, their spectra contained a strong low-frequency component...

The lax stops also show a significant drop in level in the high frequencies. This high-frequency loss is a consequence of the lower pressure associated with the production of lax stops and is therefore a crucial cue for this class of stops.”

Halle, Hughes, and Radley (1957)

Background: Production
Background: Production

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/b/</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hz</td>
<td>1910</td>
<td>1163</td>
<td>747*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>/t/</th>
<th>/d/</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hz</td>
<td>3600</td>
<td>3300</td>
<td>300*</td>
</tr>
<tr>
<td></td>
<td>5649</td>
<td>5225</td>
<td>424*</td>
</tr>
<tr>
<td></td>
<td>4900</td>
<td>4400</td>
<td>500*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>/k/</th>
<th>/g/</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hz</td>
<td>1940</td>
<td>1910</td>
<td>30*</td>
</tr>
<tr>
<td></td>
<td>2261</td>
<td>2268</td>
<td>-7*</td>
</tr>
</tbody>
</table>

+ = Zue (1976) using peak frequency
♀ = Parikh and Loizou (2005) using peak frequency
♀ = Sundara (2005) using mean frequency (CoG)

see also Van Alphen and Smits (2004), Vicenik (2010), Kirkham (2011)
production study
laboratory and TIMIT experiments
methods adapted from Forrest et al. (1988), Jongman et al. (2000), Sundara (2005)

/p,t,k,b,d,g/ x /i,i,e,ɛ,æ,ʌ,a,o,u/ x /t/

N=18 (4 male)
resampled at 16kHz
pre-emphasized above 1000Hz
high-pass filtered at 200Hz
segmented from transient to voicing

Laboratory Production: Methods
- Computed 64-point FFT for 7 consecutive 3ms Hamming windows, shifted by 1ms
- 7 PSDs averaged to give a smoothed spectrum
- Center of Gravity (CoG) calculated from smoothed spectrum: amplitude-weighted mean frequency
  \[ CoG = f_1 p(1) + \ldots + f_{32} p(32) \]

Laboratory Production: Measurement
Laboratory Production: Results

- **lab**
  - vcl: 3318 Hz
  - vcd: 2833 Hz

- **cor**
  - vcl: 4967 Hz
  - vcd: 4664 Hz

- **dor**
  - vcl: 3450 Hz
  - vcd: 3521 Hz

The symbol * indicates a statistically significant difference.
Mixed-effects linear regression
Fixed effects sum-coded and maximal random effect structure

\[ \beta_{\text{voice}} = 122, \ p < .01 \]
\[ \times \ \text{place} \] \[ \beta_{\text{labial}} = -633, \ p < .001; \beta_{\text{coronal}} = 916, \ p < .001 \]
\[ \times \ \text{gender} \] \[ \beta_{\text{gender}} = 86, \ p < .01 \]

Significant interactions examined with post-hoc comparisons

<table>
<thead>
<tr>
<th></th>
<th>labial</th>
<th>coronal</th>
<th>dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>male</strong></td>
<td>( \beta_{\text{voice}} = 224 )</td>
<td>( \beta_{\text{voice}} = 224 )</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>( p &lt; .01 )</td>
<td>( p &lt; .05 )</td>
<td></td>
</tr>
<tr>
<td><strong>female</strong></td>
<td>( \beta_{\text{voice}} = 253 )</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>( p &lt; .001 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Crucially, the pattern of significance remains the same when tokens with glottal pulses near the release are excluded.

Laboratory Production: Analysis
630 different AE speakers
Word-initial, pre-vocalic /p, t, k, b, d, g/
Words with high token freq. removed (too, to, do, carry, dark)

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>661</td>
</tr>
<tr>
<td>/t/</td>
<td>579</td>
</tr>
<tr>
<td>/k/</td>
<td>1179</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>/b/</td>
<td>668</td>
</tr>
<tr>
<td>/d/</td>
<td>547</td>
</tr>
<tr>
<td>/g/</td>
<td>415</td>
</tr>
</tbody>
</table>
### TIMIT: Results

<table>
<thead>
<tr>
<th>Voicing</th>
<th>lab</th>
<th>cor</th>
<th>dor</th>
</tr>
</thead>
<tbody>
<tr>
<td>vcl</td>
<td>3704</td>
<td>4550</td>
<td>3155</td>
</tr>
<tr>
<td>vcd</td>
<td>2672</td>
<td>3743</td>
<td>2941</td>
</tr>
</tbody>
</table>

CoG (Hz) with significant differences indicated by (*).
Mixed-effects linear regression
Fixed effects sum-coded and maximal random effect structure

\[ \beta_{\text{voice}} = 320, \ p < .001 \]
\[ \beta_{\text{labial}} = -314, \ p < .001; \beta_{\text{coronal}} = 762, \ p < .001 \]
\[ \beta_{\text{gender}} = 205, \ p < .001 \]

Significant interactions examined with post-hoc comparisons

<table>
<thead>
<tr>
<th></th>
<th>labial</th>
<th>coronal</th>
<th>dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>male</strong></td>
<td>[\beta_{\text{voice}} = 555 \text{ } p &lt; .001]</td>
<td>[\beta_{\text{voice}} = 460 \text{ } p &lt; .001]</td>
<td>(\beta_{\text{voice}} = 112 \text{ } p &lt; .001)</td>
</tr>
<tr>
<td><strong>female</strong></td>
<td>[\beta_{\text{voice}} = 396 \text{ } p &lt; .001]</td>
<td>[\beta_{\text{voice}} = 280 \text{ } p &lt; .001]</td>
<td>(\beta_{\text{voice}} = 113 \text{ } p &lt; .05)</td>
</tr>
</tbody>
</table>

Crucially, the pattern of significance remains the same, except for the dorsals, when tokens with glottal pulses near the release are excluded.

**TIMIT: Analysis**
perception study
laboratory and Mechanical Turk experiments
Trading relation between burst and VOT

/t/-burst VOT continuum  /d/-burst VOT continuum

Keating (1979)
Nittrouer (1999)
Caldwell and Nittrouer (2013)

Background: Perception
Labial Continua /bæt/-/pæt/

Keating (1979), Ganong (1980), Andruski et al. (1994)

CoG: 3494Hz
Dur: 10ms

CoG: 1513Hz
Dur: 10ms

VOT (ms)

10
17
24
31
38
45
52
Coronal Continua /dat/-/tat/

Keating (1979), Ganong (1980), Andruski et al. (1994)

Laboratory Perception: Stimuli
Two-alternative forced choice identification

Order of labial and coronal conditions counterbalanced

Within condition: 8 blocks of 14 stimuli in random order

Laboratory Perception: Methods and analysis
Laboratory Perception: Results

\[
\beta_{\text{burst}} = .54 \\
p < .001 \\
N = 16
\]
Labial Perception: Results

N=16
Laboratory Perception: Results

$\beta_{\text{burst}} = .85$

$p < .001$

$N = 16$
Laboratory Perception: Results

N=16

coronals
Crowdsourcing service increasingly used in psycholinguistics and phonetic studies

Greater diversity in participant population and listening conditions (noise!)

Labials
- 12 headphones
- 3 external speakers
- 1 internal speakers

Coronals
- 9 headphones
- 4 external speakers
- 3 internal speakers
Mechanical Turk: Results

labials

\[ \beta_{\text{burst}} = 0.46 \]
\[ p < 0.001 \]
\[ N = 16 \]
Mechanical Turk: Results

\[ \beta_{\text{burst}} = .60 \]
\[ p < .001 \]
\[ N = 16 \]
Spectral shape of the burst is a cue to anterior stop consonant voicing

Higher CoG for voiceless labials and coronals

Spectral shape influences voicing identification

Summary and Implications
Place and voice perception are interdependent

Cues to phonetic distinctions at burst landmark

Early cue to voicing and incremental perception

Summary and Implications
Thank you!
Production: Results by Gender

- CoG (Hz)
- TIMIT
- laboratory
Mechanical Turk: Results

N=16

labials
Mechanical Turk: Results

N=16 coronals
<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Measure</th>
<th>/p/</th>
<th>/b/</th>
<th>/t/</th>
<th>/d/</th>
<th>/k/</th>
<th>/g/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zue 1976</td>
<td>Am. English</td>
<td>Peak</td>
<td>--</td>
<td>--</td>
<td>3600</td>
<td>3300</td>
<td>1940</td>
<td>1910</td>
</tr>
<tr>
<td>Parikh and Loizou 2005</td>
<td>Am. English</td>
<td>Peak</td>
<td>1910</td>
<td>1163</td>
<td>5649</td>
<td>5225</td>
<td>2261</td>
<td>2268</td>
</tr>
<tr>
<td>Sundara 2005</td>
<td>Ca. English</td>
<td>CoG</td>
<td>--</td>
<td>--</td>
<td>4900</td>
<td>4400</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Kirkham 2011</td>
<td>Br. English</td>
<td>CoG</td>
<td>--</td>
<td>--</td>
<td>5220</td>
<td>4888</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Van Alphen and Smits 2004</td>
<td>Dutch</td>
<td>CoG</td>
<td>1160</td>
<td>830</td>
<td>3540</td>
<td>2140</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Sundara 2005</td>
<td>Ca. French</td>
<td>CoG</td>
<td>--</td>
<td>--</td>
<td>3800</td>
<td>3000</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Vicenik 2010</td>
<td>Georgian</td>
<td>CoG</td>
<td>4000</td>
<td>3200</td>
<td>5300</td>
<td>4600</td>
<td>3100</td>
<td>3100</td>
</tr>
</tbody>
</table>

CoG = Center of Gravity (mean frequency)