A new metric for calculating acoustic dispersion in stop inventories

Introduction

- Dispersion Theory (DT; Liljencrants and Lindblom, 1972) claims acoustically dispersed vowel inventories should be typologically common.
- Dispersion is often quantified using triangle area between three mean or prototypical points (Andruski et al., 1999; Jacewicz et al., 2007; Ziegler & Von Cramon, 1983).
  → This approach ignores distributions, reducing variable categories to single points.
  → Within-category variance affects perception (Clayards, 2008).
- **Proposal**: A new metric that incorporates within-category spread based on the Jeffries-Matusita (JM) distance.
- **Results**: When using the new metric to extend DT predictions to consonants, results change. But still does not recover predictions of DT for stops.
  
  **Choice of metric is important and influences results in work on Dispersion Theory.**

Dispersion Theory and consonants

- There has been relatively little work on whether DT also applies to consonant inventories.
  
  → Schwartz et al. (2012). The /coronal-velar-pharyngeal/ (/d g/)) inventory is the most dispersed, not the typologically common /labial-coronal-velar/ (/b d g/).

Hypotheses to understand this result in light of DT predictions:

(1) DT does not apply to consonants.
(2) The phonetic space they considered is not the most relevant for the data
   → Argued by Schwartz et al. using a Frame Content model (MacNeille, 1998) to exclude pharyngeals and epiglottals.
(3) The metric they chose is not the most relevant for the data (proposed here).

- Even with the more complex metric shown here, the /labial-coronal-velar/ (/b d g/) inventory is never among the most dispersed.

**Implications**: DT does not apply to stop inventories the same way it applies to vowel inventories.

- DT either makes incorrect predictions for stop inventories, or an approach like that of Schwartz et al. is needed.

The data (Schwartz et al., 2012)

- 50,000 stop tokens generated along vocal tract in three vowel contexts ([i a u]), then grouped according to 8 places of articulation (POA).

F1, F2, F3 sampled at beginning of transition to vowel when formant structure first appears.

- This is a reasonable phonetic space to consider for stop POA.
- Evidence points to primacy of formant transitions as perceptual cues to POA (Walley & Carrell, 1988; Sussman et al., 1991).

Problems characterizing distributions with means

- Shapes and variances of categories are lost when characterized only by mean values.

Calculating dispersion with mean-to-mean distance

\[
\text{Distance between two mean points (i, j) in } < F_1, F_2, F_3 > \text{ space} \]

\[
d_{ij} = \sqrt{(F_1_i - F_1_j)^2 + (F_2_i - F_2_j)^2 + (F_3_i - F_3_j)^2} \quad (1)
\]

Area of a triangle as a dispersion metric

\[
A = \frac{1}{2} |(s - d_a)(s - d_b) - s(s - d_a + d_b)| \quad \text{where } s = (d_a + d_b + s)/3 \quad (2)
\]

**Mean-to-mean dispersion results in < F_1, F_2, F_3 > space**

<table>
<thead>
<tr>
<th>POA1</th>
<th>POA2</th>
<th>POA3</th>
<th>Dispersion (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 coronal</td>
<td>epi-pharyngeal velar</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>2 coronal</td>
<td>uvular velar</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>3 epi-pharyngeal palatal</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 bilabial coronal epi-pharyngeal</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 bilabial coronal velar</td>
<td>0.23 = Expected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Incorporating covariance into dispersion

- Jeffries-Matusita (JM) distance (Kobayashi & Thomas, 1967) used as the base of a dispersion metric, incorporates covariance, the multidimensional analog of variance.
  → A transformation of the Bhattacharyya distance into a fixed range $[0, \sqrt{2}]$.

**Jeffries-Matusita Distance between two distributions F,G**

\[
D_{JM}(F, G) = \sqrt{2(1 - \exp(-D_{JM}(F, G)))} \quad \text{where } D_{JM}(F, G) = \int \sqrt{f(x)g(x)}dx \quad (3)
\]

**JM distance dispersion results in < F_1, F_2, F_3 > space**

<table>
<thead>
<tr>
<th>POA1</th>
<th>POA2</th>
<th>POA3</th>
<th>Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 coronal</td>
<td>epi-pharyngeal velar</td>
<td>0.864</td>
<td></td>
</tr>
<tr>
<td>2 coronal</td>
<td>epi-pharyngeal palatal</td>
<td>0.863</td>
<td></td>
</tr>
<tr>
<td>15 coronal palatal</td>
<td>0.781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 coronal velar</td>
<td>0.773 = Expected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Jeffries-Matusita distance does not improve the results in favor of DT predictions.**

- The desired /b d g/ inventory ranks lower with this metric than mean-to-mean distance.
- Epi-pharyngeals will always be present in the most dispersed inventories.

Discussion and conclusion

Is < F_1, F_2, F_3 > the relevant space for this data?

- Maybe the F1 dimension isn’t as relevant in consonant perception. F1 only discriminates epi-pharyngeals from other places.

**JM dispersion results in < F_2, F_3 > space**

<table>
<thead>
<tr>
<th>POA1</th>
<th>POA2</th>
<th>POA3</th>
<th>Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 coronal</td>
<td>epi-pharyngeal velar</td>
<td>0.836</td>
<td></td>
</tr>
<tr>
<td>2 coronal</td>
<td>epi-pharyngeal palatal</td>
<td>0.774</td>
<td></td>
</tr>
<tr>
<td>10 coronal palatal</td>
<td>0.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 coronal velar</td>
<td>0.697 = Expected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**But even when removing the F1 dimension entirely, the /b d g/ inventory is still not among the most dispersed, and the better dispersed inventories are all unattested.**

Sufficient vs. maximal dispersion (Lindblom, 1986)

- The /b d g/ inventory can’t be considered sufficiently dispersed while /d g/ is maximally dispersed because inventories with similar dispersion scores to /d g/ are unattested.

Conclusion

Results always depend on the metric chosen in work on Dispersion Theory.

- The most acoustically dispersed stop inventory is unattested ([/d g/]), not the typologically common /b d g/.
- Either Dispersion Theory doesn’t apply to consonants as it does vowels, or the phonetic space must be altered to exclude pharyngeals and epiglottals as in Schwartz et al. (2012).