Finding phonological structure in vowel confusions across English accents

Jason Shaw, Catherine Best, Paul Foulkes, Bronwen Evans, Gerard Docherty, Karen Mulak

Berkeley Linguistics Society Workshop
Phonological Representations: At the Crossroad Between Gradience and Categoricity
Feb 7-8, 2020
Outline

1) Present a series of cross-accent vowel categorization studies
   – Listeners from five (non-rhotic) English accents categorized their native accent vowels.
   – Listeners from one of those accents also categorized vowels of the four other (non-rhotic) English accents

2) Train and test computational models on the same tasks.

3) Argue that contrastive feature hierarchies provide particular insight into confusion patterns within and across accents.
Assumption

• *(We think)* there is broad agreement that speech is perceived in terms of phonological representations, whatever they might be (e.g., Goldinger 1998; Fowler 1986; Poeppel, Idsardi, van Wassenhove 2008).

• Make that assumption here:
  phonological representation \(\approx\) object of speech perception
What factors dictate perceptual confusion?

$H_0$: **acoustic distinctiveness**—acoustically similar sounds get confused

$H_1$: **phonological distinctiveness**—phonologically similar sounds will get confused

Within a speech community, phonetics and phonology co-evolve such that it may be difficult to test $H_1$ (phonological contrasts tend to be robust in the acoustics)

*Cross-accent perception offers an opportunity to dissociate acoustic distinctiveness and phonological distinctiveness*
Test case: cross-accent perception

- Non-rhotic English accents have similar numbers of vowels but they differ in their **phonetic realizations** and **corresponding phonological structure**, e.g., expressed in terms of contrastive feature hierarchies (e.g., Dresher 2009)

- **Successive Division Algorithm** (Dresher 2009: 16)
  a. Begin with *no* features specifications: assume all sounds are allophones of a single undifferentiated phoneme.
  b. If the set is found to consist of more than one contrasting member, select a features and divide the set into as many subsets as the feature allows for.
  c. Repeat previous step in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.
Partial feature hierarchies for vowels of Australian (left) and New Zealand (right) English

A similar inventory of vowels can have different contrastive feature hierarchies
Key comparisons: native & cross-accent perception

Aussie baseline
- Aussie listeners
- Aussie vowels

Similarity based on **imposition of listener phonology (feature hierarchy)**

NZ baseline
- New Zealand listeners
- New Zealand vowels

**Prediction:** if confusions are structured by a feature hierarchy ... **cross-accent confusions should resemble Aussie baseline more than confusions across accents**

Test condition
- Aussie listeners
- New Zealand vowels

Differences in perceptual structure between accents (due to phonetic and phonological differences)
EXPERIMENTAL METHOD

STIMULI
- 5 accents: Australia, London, New Zealand, Yorkshire, Newcastle
- Nonce words for the 20 English (lexical set) vowels produced in /zVba/ frame by 4 speakers (2 ♀, 2 ♂) per accent (used 2 tokens/speaker x 2 reps/token)

LISTENERS
- 9 conditions: Aussies heard all 5 accents; other groups heard own accent only
- 12-16 monolingual listeners per condition (136 total; M_{age} = 22)

VOWEL CATEGORIZATION TASK
- Categorized nonce words to 19 keywords

ANALYSES
1. Cross-accent confusions
2. Human vs. machine
Cluster analysis of a whole-system confusion matrix

- Confusion matrices were progressively fused into binary clusters that minimize the variance of each cluster (Ward’s method)
- Resulting hierarchical clusters represent the perceptual structure imposed on the stimuli by listeners
Tanglegram comparing structures

- Tanglegrams illustrate differences between 2 hierarchical clusters

- Baker’s Gamma
  - Correlation coefficient between the 2 clustered objects
  - Quantitative measure of similarity (0 to 1); 1 = perfect match

Baker’s Gamma = 0.44
RESULTS
Cluster analyses: Native Accents comparison

Australian listeners on Australian vowels

New Zealand listeners on New Zealand vowels

Baker’s Gamma = 0.35
Cluster analyses: Australian listeners on both accents

Australian listeners on Australian vowels

Australian listeners on New Zealand vowels

Baker’s Gamma = 0.44
Baker’s Gamma comparisons

<table>
<thead>
<tr>
<th>Baker’s Gamma</th>
<th>Vowel accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>New Zealand</td>
</tr>
<tr>
<td>0.38</td>
<td>Newcastle</td>
</tr>
<tr>
<td>0.35</td>
<td>London</td>
</tr>
<tr>
<td>0.34</td>
<td>Yorkshire</td>
</tr>
</tbody>
</table>

CONTROL:
Australians on AusE re: native listeners of each other accent

<table>
<thead>
<tr>
<th>Baker’s Gamma</th>
<th>Vowel accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.44</td>
<td>New Zealand</td>
</tr>
<tr>
<td>0.45</td>
<td>Newcastle</td>
</tr>
<tr>
<td>0.45</td>
<td>London</td>
</tr>
<tr>
<td>0.47</td>
<td>Yorkshire</td>
</tr>
</tbody>
</table>

less similar than

CROSS-ACCENT TEST:
Australian listeners on each non-AusE accent’s vowels

Always < 1, indicating different perceptual structures across listeners’ accents

Aussie perceptual structure imposed on other accent, evidence for **perceptual assimilation** across accents
Are confusions due to acoustic similarity alone? human vs. machine approach

**HUMAN**

- Auditory stimuli ➔ Categorization task ➔ Confusion matrix ➔ Hierarchical structure

**MACHINE**

- Auditory stimuli ➔ Categorization task ➔ Confusion matrix ➔ Hierarchical structure

Does bottom-up classification of the signal result in the human pattern?
Computational Method

• Multinomial Logistic Regression, following McMurray & Jongman’s (2011) work on English fricatives.

• Compared several sets of acoustic features for vowels
  – $F1(50\%) + F2(50\%) + \text{(duration)}$
  – $F1(20\%) + F1(80\%) + F2(20\%) + F2(80\%) + \text{(duration)}$
  – First two DCT coefficients fit to change in $F1$ and $F2$ across the vowel + (duration)
  – First 2-5 Principal Components of MFCCs + (duration)

• Evaluated models based on:
  – Variance explained given complexity: Akaike Information Criteria (AIC)
  – Correspondence with perceptual data: Baker’s Gamma
Acoustic parameter comparison: AIC

<table>
<thead>
<tr>
<th>AIC</th>
<th>F1 + F2</th>
<th>F1 + F2 + Dur</th>
<th>F1.20 + F1.80 + F2.20 + F2.80 + duration</th>
<th>MFCC PC1-5</th>
<th>MFCC PC1+ PC2</th>
<th>MFCC PC1+ PC2+ duration</th>
<th>2DCT F1 + F2</th>
<th>2DCT F1 + F2 + duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian</td>
<td>556</td>
<td>337</td>
<td>338</td>
<td>283</td>
<td>798</td>
<td>616</td>
<td>430</td>
<td>298</td>
</tr>
<tr>
<td>London</td>
<td>601</td>
<td>421</td>
<td>409</td>
<td>321</td>
<td>1024</td>
<td>642</td>
<td>498</td>
<td>378</td>
</tr>
<tr>
<td>New Zealand</td>
<td>439</td>
<td>357</td>
<td>297</td>
<td>284</td>
<td>1019</td>
<td>767</td>
<td>568</td>
<td>283</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>454</td>
<td>309</td>
<td>388</td>
<td>302</td>
<td>952</td>
<td>474</td>
<td>321</td>
<td>365</td>
</tr>
<tr>
<td>Newcastle</td>
<td>558</td>
<td>400</td>
<td>293</td>
<td>225</td>
<td>918</td>
<td>667</td>
<td>529</td>
<td>302</td>
</tr>
</tbody>
</table>

*Lower numbers indicate better model*
Acoustic parameter comparison: Baker’s Gamma with human data

<table>
<thead>
<tr>
<th>Baker’s Gamma</th>
<th>F1 + F2</th>
<th>F1 + F2 + Dur</th>
<th>F1.20 + F1.80 + F2.20 + F2.80 + duration</th>
<th>MFCC PC1-5</th>
<th>MFCC PC1+ PC2</th>
<th>2DCT F1 + F2</th>
<th>2DCT F1 + F2 + duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>0.08</td>
<td><strong>0.17</strong></td>
<td>0.01</td>
<td>0.07</td>
<td>0.04</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>London</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.21</td>
<td>0.26</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.29</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.32</td>
<td>0.16</td>
<td>0.06</td>
<td>0.09</td>
<td><strong>0.38</strong></td>
<td>0.37</td>
<td>0.15</td>
</tr>
<tr>
<td>Yorkshire</td>
<td>0.14</td>
<td>0.21</td>
<td>0.26</td>
<td>0.35</td>
<td>0.16</td>
<td>0.37</td>
<td>0.24</td>
</tr>
<tr>
<td>Newcastle</td>
<td>0.10</td>
<td>-0.05</td>
<td>0.06</td>
<td>0.29</td>
<td>-0.05</td>
<td>-0.00</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*higher numbers indicate closer approximation to human patterns*
Baker’s Gamma comparisons: machine

**CONTROL:**
Australian training data on Australian stimuli re: each other accent

<table>
<thead>
<tr>
<th>Baker’s Gamma</th>
<th>Vowel accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.23</td>
<td>New Zealand</td>
</tr>
<tr>
<td>-0.15</td>
<td>Newcastle</td>
</tr>
<tr>
<td>0.46</td>
<td>London</td>
</tr>
<tr>
<td>0.13</td>
<td>Yorkshire</td>
</tr>
</tbody>
</table>

**CROSS-ACCENT TEST:**
Australian training data tested on each non-AusE accent’s vowels

<table>
<thead>
<tr>
<th>Baker’s Gamma</th>
<th>Vowel accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.09</td>
<td>New Zealand</td>
</tr>
<tr>
<td>0.26</td>
<td>Newcastle</td>
</tr>
<tr>
<td>0.26</td>
<td>London</td>
</tr>
<tr>
<td>0.03</td>
<td>Yorkshire</td>
</tr>
</tbody>
</table>

*In contrast to human data, no clear relation between control (left) and cross-accent test (right)*
Summary

• Patterns of *cross-accent* perceptual confusion more closely resemble listeners' confusions in their *native accent* than confusions based on the other unfamiliar accent.

• This pattern can’t be derived (so far, anyway) from bottom-up acoustic similarity.
Discussion

• We know that phonology shapes perception:
  – Perceptual Assimilation in cross-language speech (e.g., Meinhoff 1933)
  – Perceptual “illusions” conditioned by phonotactics, syllable structure, lexical stress, phonological rules, phonological phrasing, etc.
  – All point to a crucial role for phonological expectations (priors) in perception

• Results here indicate that listeners impose native accent perceptual structure on unfamiliar accents.

• Contrastive feature hierarchies have the potential to account for differences across accents and in cross-accent perception (for mathematical basis in information theory see: Shaw et al 2019)
Funding acknowledgment

Research supported by:

- ARC grant: DP120104596
Vowel formants of five non-rhotic accents of English