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

Yale
WESTERN SYDNEY UNIVERSITY
Griffith UNIVERSITY
UCL
UNIVERSITY of York
Newcastle University

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 ≈ 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

[coronal]

- +

ɒ NORTH (BORED)

ɐ STRUT (BUD)

ʉ GOOSE (RUDE)

[low]

- +

[high]

- +

æ TRAP (BAD)

e DRESS (BED)

ɪ KIT (BID)

[round]

- +

[coronal]

- +

ʉ GOOSE (RUDE)

ɒ NORTH (BORED)

[low]

- +

ɪ DRESS (BED)

ɛ TRAP (BAD)

ɘ KIT (BID)

ɐ STRUT (BUD)

A similar inventory of vowels can have different contrastive feature hierarchies

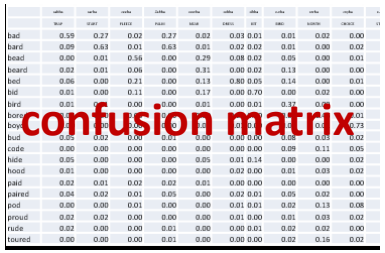
Key comparisons: native & cross-accent perception

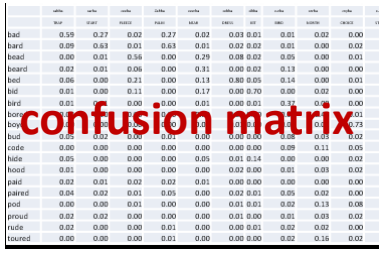
Aussie baseline

Aussie listeners

Aussie vowels

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





Test condition

Aussie listeners

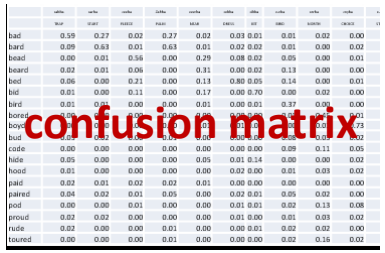
New Zealand vowels

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

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

EXPERIMENTAL METHOD

STIMULI

- 5 accents: Australia, London, New Zealand, Yorkshire, Newcastle
- **Nonce** words for the 20 English (lexical set) vowels produced in **/zVbə/** 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

bard	beard	boyd	paired	bad
rude	bead	bored	pod	code
hood	bud	bid	hide	bed
proud	toured	bird	paid	

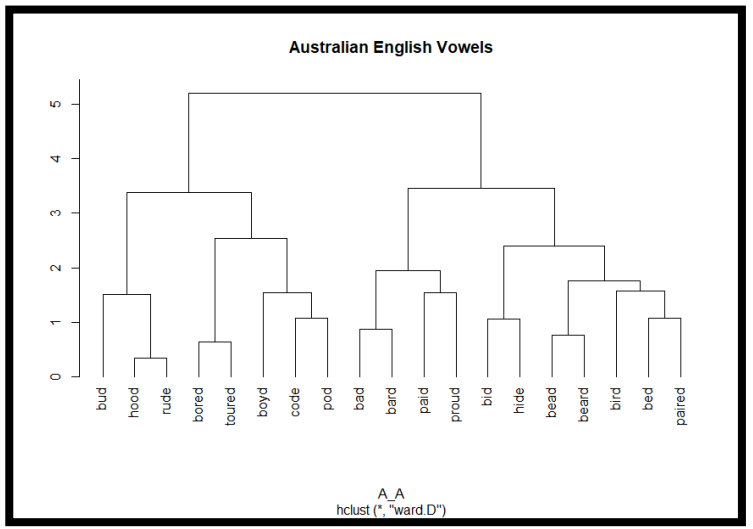
ANALYSES

1. Cross-accent confusions
2. Human vs. machine

Cluster analysis of a whole-system confusion matrix

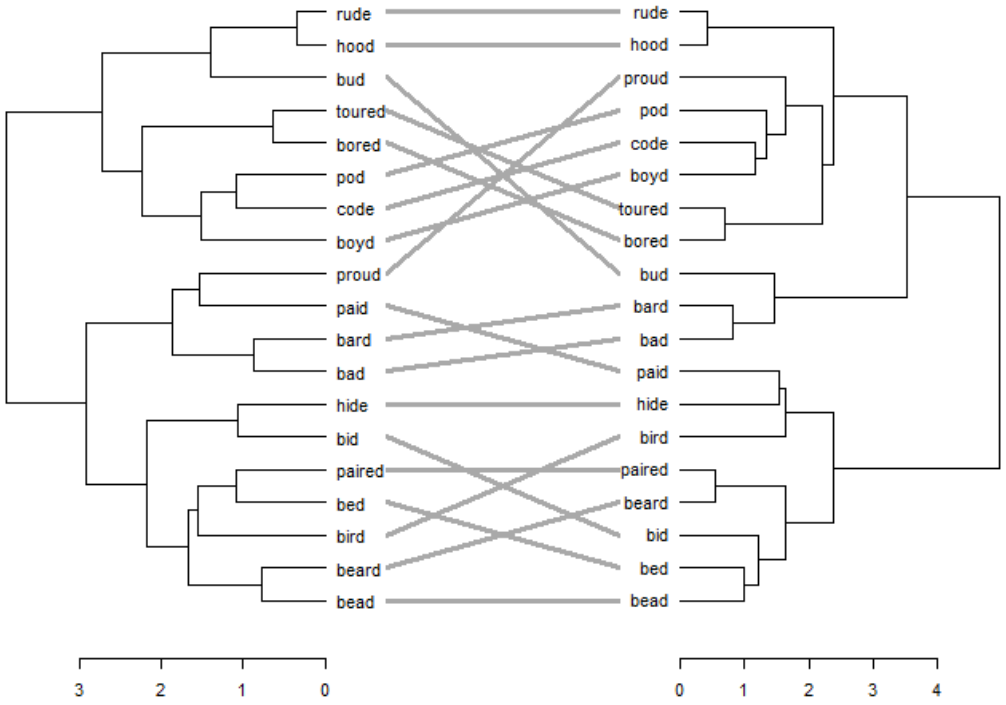
	zabba	zarba	zeeba	Zahba	zeerba	zebba	zibba	zuba	zorba	zoyba	zubba	zoeba	zieba	zooba	zayba	zairba	zobba	zowba	zewba	zourba
	TRAP	START	FLEECE	PALM	NEAR	DRESS	KIT	BIRD	NORTH	CHOICE	STRUT	GOAT	PRICE	FOOT	FACE	SQUARE	LOT	MOUTH	GOOSE	TOUR
bad	0.59	0.27	0.02	0.27	0.02	0.03	0.01	0.01	0.02	0.00	0.34	0.00	0.01	0.00	0.31	0.04	0.02	0.20	0.00	0.01
bard	0.09	0.63	0.01	0.63	0.01	0.02	0.02	0.01	0.00	0.02	0									
bead	0.00	0.01	0.56	0.00	0.29	0.08	0.02	0.05	0.00	0.01	0									
beard	0.02	0.01	0.06	0.00	0.31	0.00	0.02	0.13	0.00	0.00	0									
bed	0.06	0.00	0.21	0.00	0.13	0.80	0.05	0.14	0.00	0.01	0									
bid	0.01	0.00	0.11	0.00	0.17	0.00	0.70	0.00	0.02	0.00	0									
bird	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.37	0.00	0.00	0									
bored	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.45	0.01	0									
boyd	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.73	0									
bud	0.05	0.02	0.00	0.01	0.00	0.00	0.00	0.08	0.03	0.02	0									
code	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.11	0.05	0									
hide	0.05	0.00	0.00	0.00	0.05	0.01	0.14	0.00	0.00	0.02	0									
hood	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.03	0.02	0									
paid	0.02	0.01	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0									
paired	0.04	0.02	0.01	0.05	0.00	0.02	0.01	0.05	0.02	0.00	0									
pod	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.02	0.13	0.08	0									
proud	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.03	0.02	0									
rude	0.02	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.02	0.00	0									
toured	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.16	0.02	0									

	zabba	zarba	zeeba
	TRAP	START	FLEECE
bad	0.59	0.27	0.02
bard	0.09	0.63	0.01
bead	0.00	0.01	0.56
beard	0.02	0.01	0.06
bed	0.06	0.00	0.21
bid	0.01	0.00	0.11



- 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



Baker's Gamma = 0.44

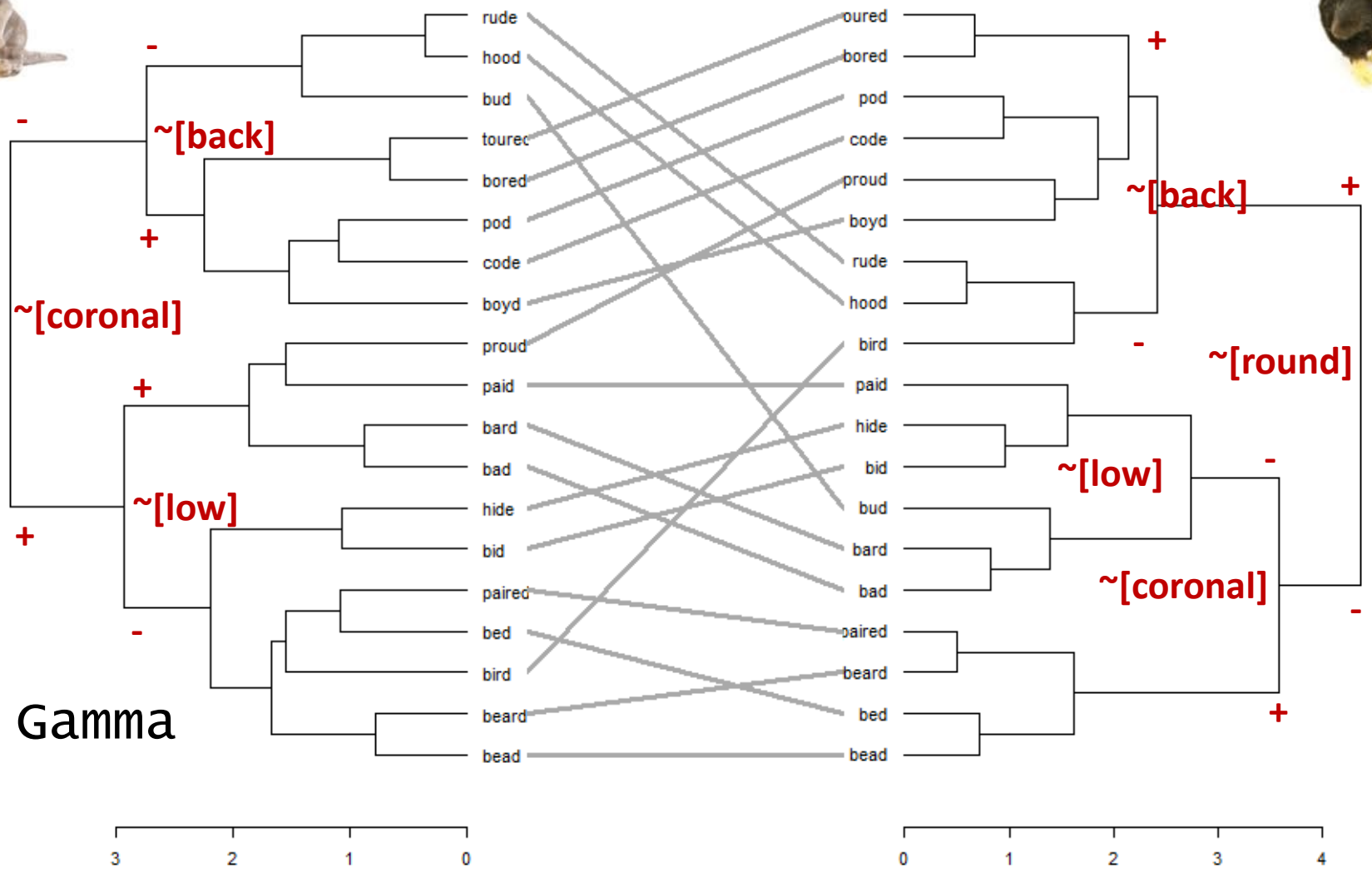
- Tanglegrams illustrate differences between 2 hierarchical clusters
- Baker's Gamma
  - Correlation coefficient between the 2 clustered objects
  - Quantitative measure of **similarity** (0 to 1); 1 = perfect match

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

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

Baker's Gamma	**Vowel accent**
0.35	New Zealand
0.38	Newcastle
0.35	London
0.34	Yorkshire

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

<
less similar than

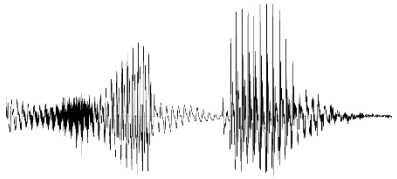
CROSS-ACCENT TEST:
Australian listeners on each non-AusE accent's vowels

Baker's Gamma	**Vowel accent**
0.44	New Zealand
0.45	Newcastle
0.45	London
0.47	Yorkshire

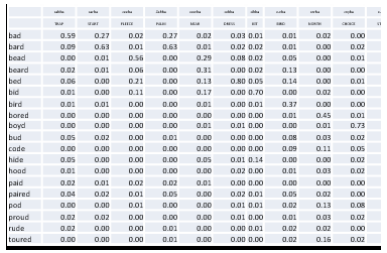
Aussie perceptual structure imposed on other accent, evidence for ***perceptual assimilation*** across accents

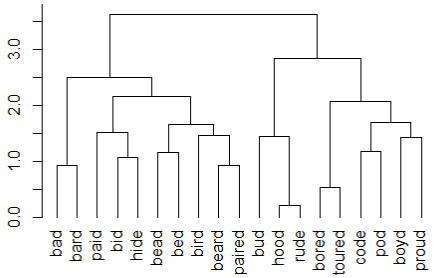
Are confusions due to acoustic similarity alone? human vs. machine approach

HUMAN



bard	beard	boyd	paired	bad
rude	bead	bored	pod	code
hood	bud	bid	hide	bed
proud	toured	bird	paid	

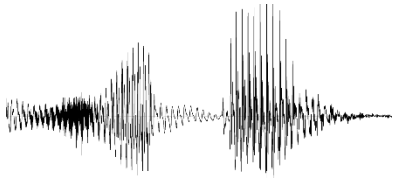




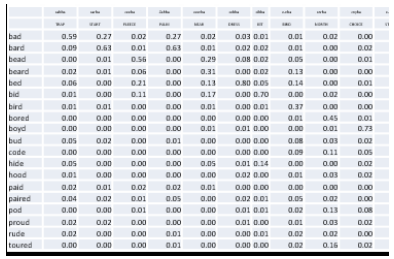
auditory stimuli → **categorization task →** **confusion matrix →** **hierarchical structure**

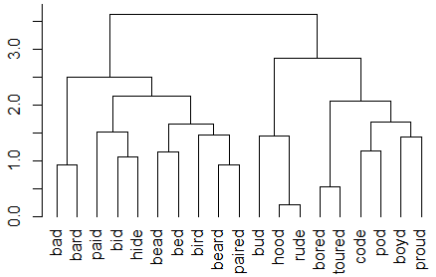
Does bottom-up classification of the signal result in the human pattern?

MACHINE



Likelihood function (MLR)





auditory stimuli → **categorization task →** **confusion matrix →** **hierarchical structure**

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%)+(duration)
  - *F1*(20%) + *F1*(80%) + *F2*(20%) + *F2*(80%)+(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

AIC	F1 + F2	F1 + F2 + Dur	F1.20 + F1.80 + F2.20 + F2.80	F1.20 + F1.80 + F2.20 + F2.80 + duration	MFCC PC1-5	MFCC PC1+ PC2	MFCC PC1+ PC2+ duration	2DCT F1 + F2	2DCT F1 + F2 + duration
Australian	556	337	338	283	798	616	430	298	**257**
London	601	421	409	321	1024	642	498	378	**305**
New Zealand	439	357	297	284	1019	767	568	283	**274**
Yorkshire	454	309	388	302	952	474	321	365	**292**
Newcastle	558	400	293	**225**	918	667	529	302	266

Lower numbers indicate better model

Acoustic parameter comparison: Baker's Gamma with human data

Baker's Gamma	F1 + F2	F1 + F2 + Dur	F1.20 + F1.80 + F2.20 + F2.80	F1.20 + F1.80 + F2.20 + F2.80 + duration	MFCC PC1-5	MFCC PC1+ PC2	2DCT F1 + F2	2DCT F1 + F2 + duration
Australian	0.08	**0.17**	0.01	0.07	0.04	0.02	0.03	0.11
London	-0.01	-0.00	0.21	0.26	-0.05	-0.05	0.29	**0.60**
New Zealand	0.32	0.16	0.06	0.09	**0.38**	0.37	0.15	0.05
Yorkshire	0.14	0.21	0.26	0.35	0.16	0.37	0.24	**0.83**
Newcastle	0.10	-0.05	0.06	0.29	-0.05	-0.00	0.06	**0.10**

higher numbers indicate closer approximation to human patterns

Baker's Gamma comparisons: machine

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

Baker's Gamma	**Vowel accent**
0.23	New Zealand
-0.15	Newcastle
0.46	London
0.13	Yorkshire

?

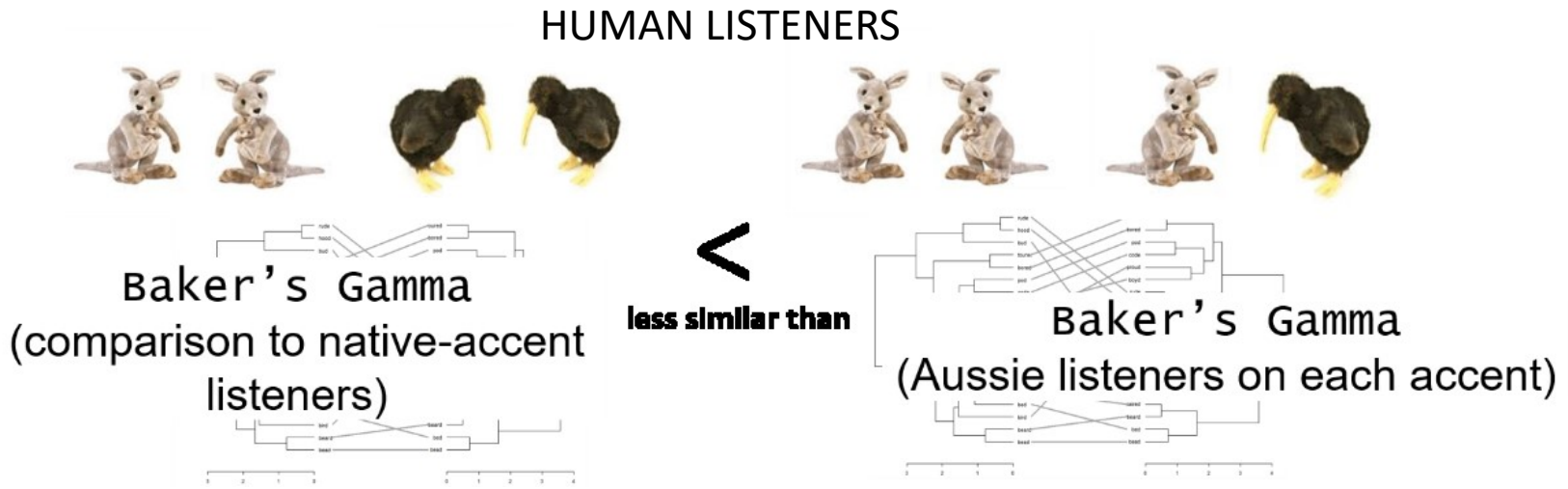
CROSS-ACCENT TEST:
Australian training data tested on each non-AusE accent's vowels

Baker's Gamma	**Vowel accent**
0.09	New Zealand
0.26	Newcastle
0.26	London
0.03	Yorkshire

*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

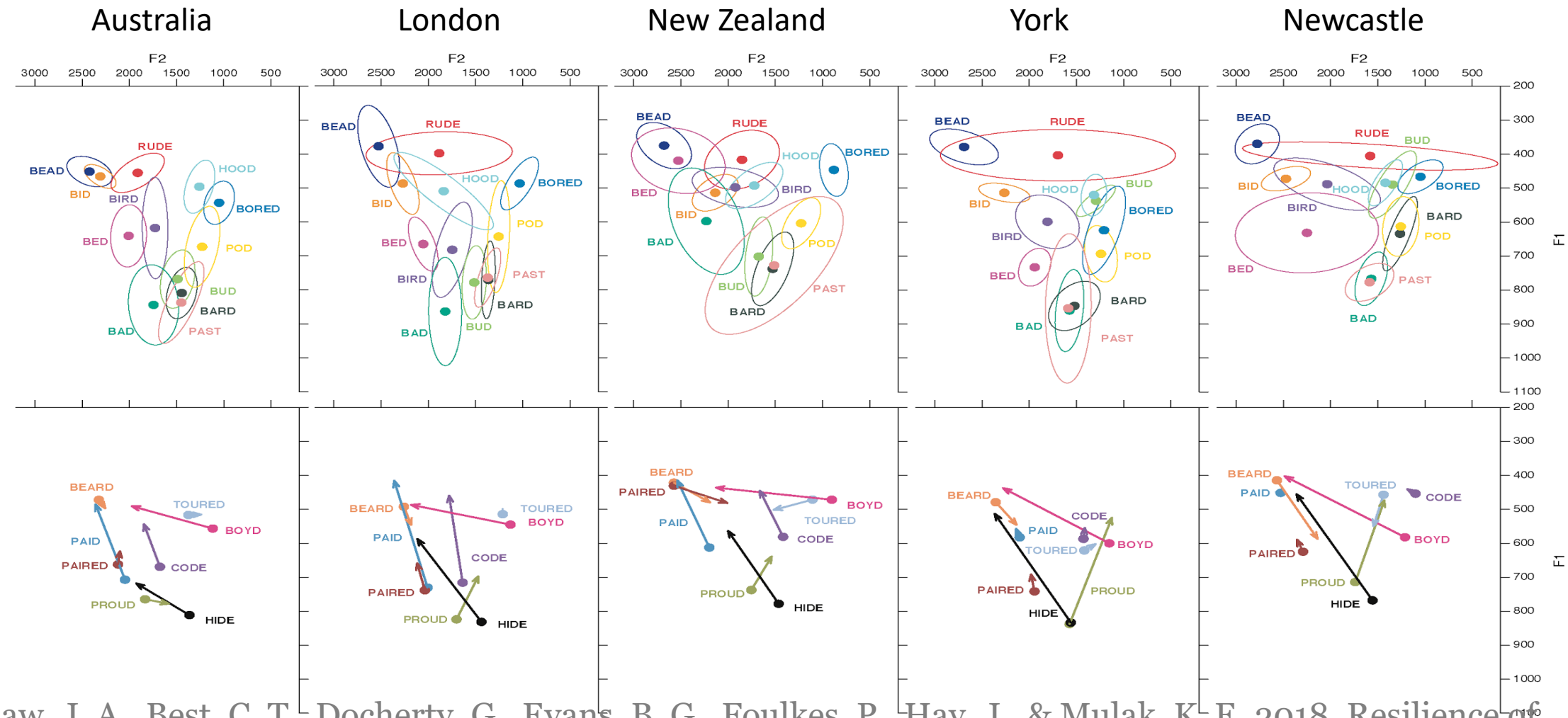
Research supported by:

- ARC grant: DP120104596

Australian Government
Australian Research Council

WESTERN SYDNEY UNIVERSITY
SCHOOL OF HUMANITIES & COMMUNICATION ARTS
Griffith UNIVERSITY
UCL
UNIVERSITY of York
UC UNIVERSITY OF CANTERBURY
Newcastle University

Vowel formants of five non-rhotic accents of English



Shaw, J. A., Best, C. T., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., & Mulak, K. E. 2018. Resilience of English vowel perception across regional accent variation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*,*9*(1), 11.