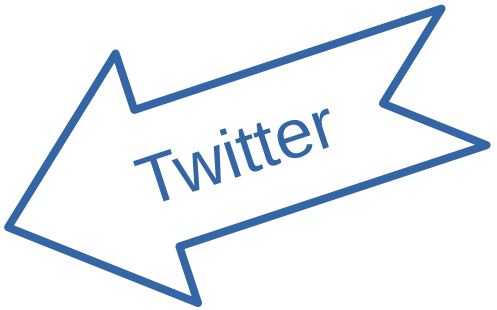
Listening with American Ears:
Using Social Information in Perceptual Learning

Rachael Tatman
University of Washington

Listening with American Ears:
Using Social Information in Perceptual Learning

Rachael Tatman
University of Washington



Dialect Perceptual Learning

- Exposure to dialect lexical items containing ambiguous sounds leads to shifting of phonemic boundary
  - Norris, McQueen & Cutler 2003
- Effect is generalizable, carries over multiple talkers
  - Kraljic & Samuel 2006
- Happens very quickly—need as few as two to four examples
  - Clarke & Garrett 2004

Dialect Perceptual Learning

- Fast, robust, transferable learning... but sometimes it doesn't transfer
- Learning on one speaker didn't carry over to a new speaker unless the original segments were spliced in
  - Eisner & McQueen 2005
- Training on a talker speaking German did not improve recognition of that same talker speaking English
  - Levi, Winters & Pisoni 2011
- So what's going on?

What's Going On?

- Listeners are (sometimes inconveniently) smart
- They can learn/extend new dialect forms
- BUT they can also “undo”/ignore variation they think is idiosyncratic
  - e.g. speaker has a pen in their mouth (Kraljic, Samuel & Brennan 2008)
- We should be able to explicitly control which tack they take by changing what social information they're exposed to

Social Information During Perception

- Social knowledge about speaker (e.g. gender) shifts expectations and percepts
  - Strand 1999
- Very little social information is needed to shift listener's expectations:
  - Note on top of questionnaire (Niedzielski 1999)
  - Socially informative stuffed animal in room (Drager & Hay 2010)

Big question:

Can the social information listeners are given control whether or not they apply recent perceptual learning?

Two possibilities:

- Social information doesn't matter, listeners rely on acoustics
- Social information does matter, can override acoustics

Two possibilities:

- Social information doesn't matter, listeners rely on acoustics
- Social information does matter, can override acoustics

What about social information in the acoustic signal?

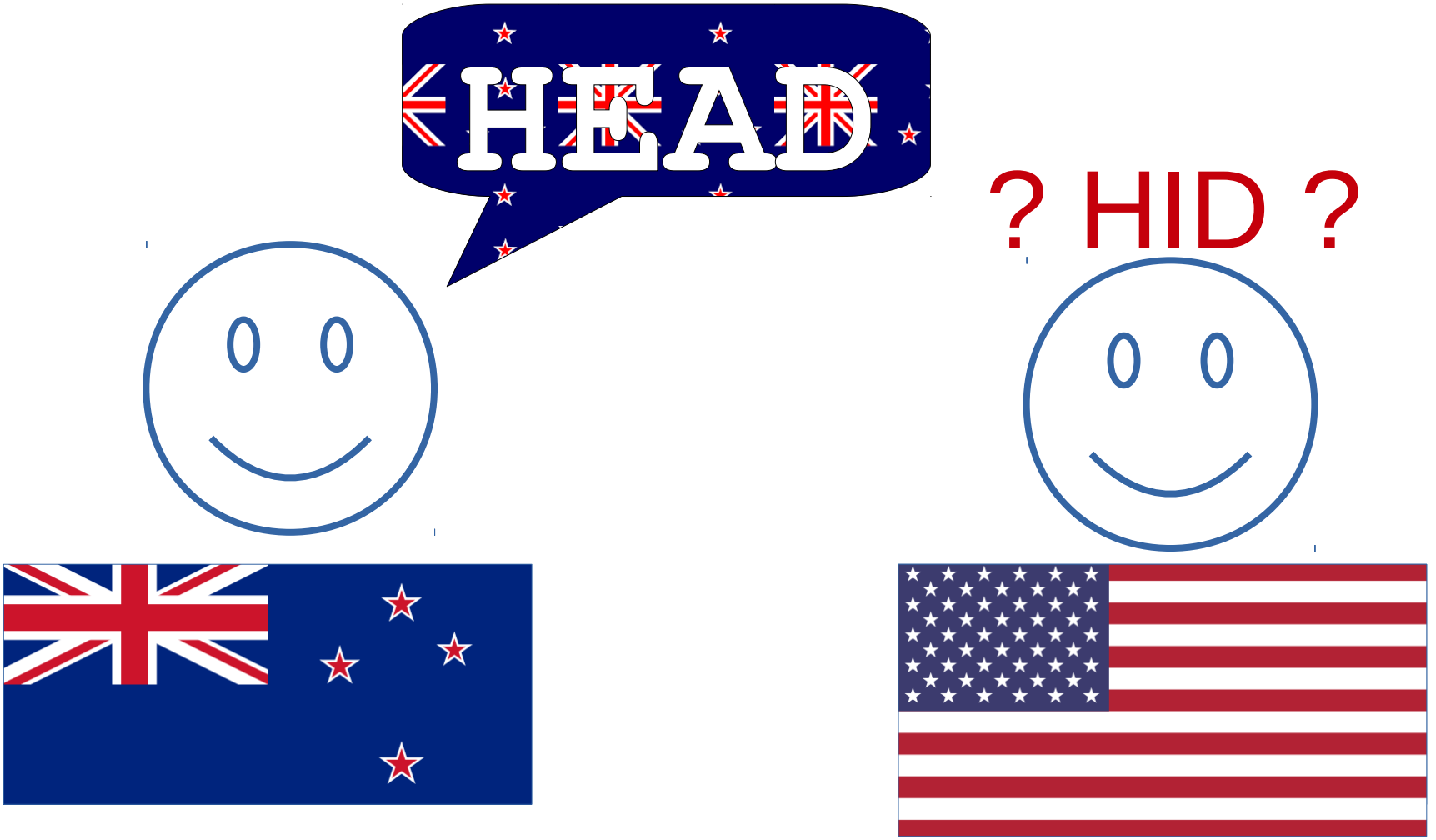
- Tried to minimize this
- Focus of further research

But how does social information effect recent perceptual learning?

- Possibilities:
  - Social information doesn't matter, listeners rely on acoustics
  - Social information does matter, can override acoustics
- What about social information in the acoustic signal?
  - Tried to minimize this
  - Focus of further research

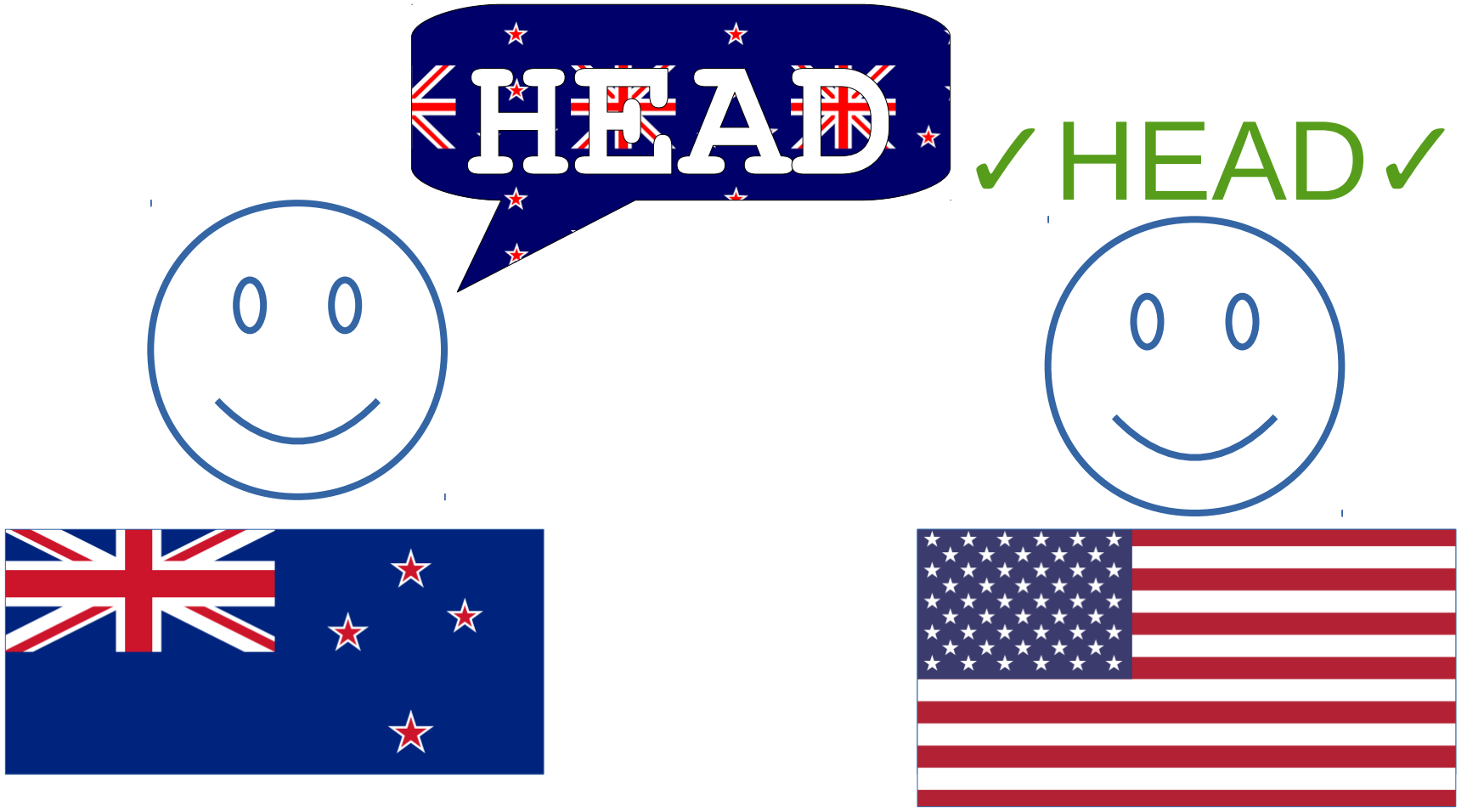
Methodology

Methodology (Overview)



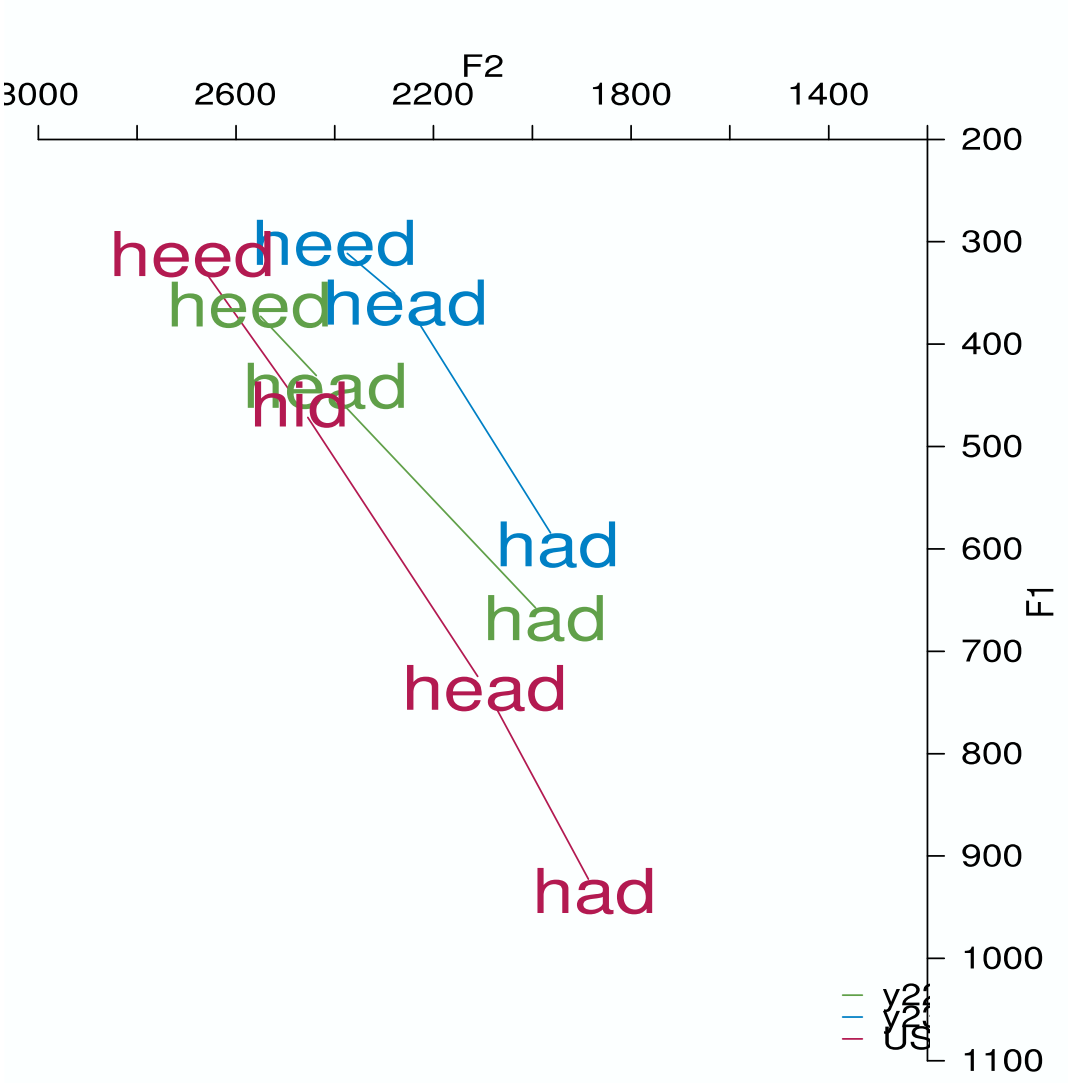
Methodology (Overview)

Methodology (Overview)

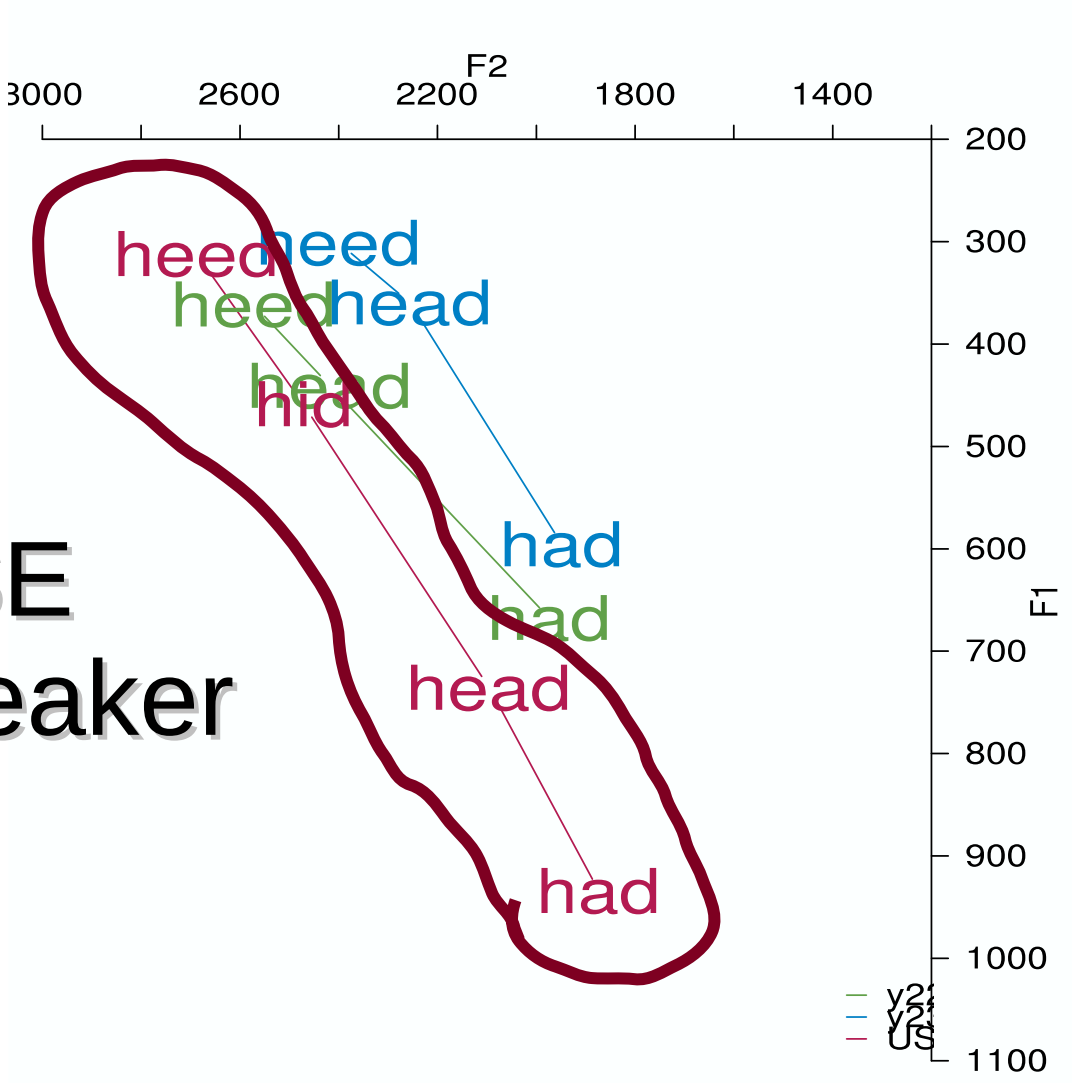


Methodology (Overview)

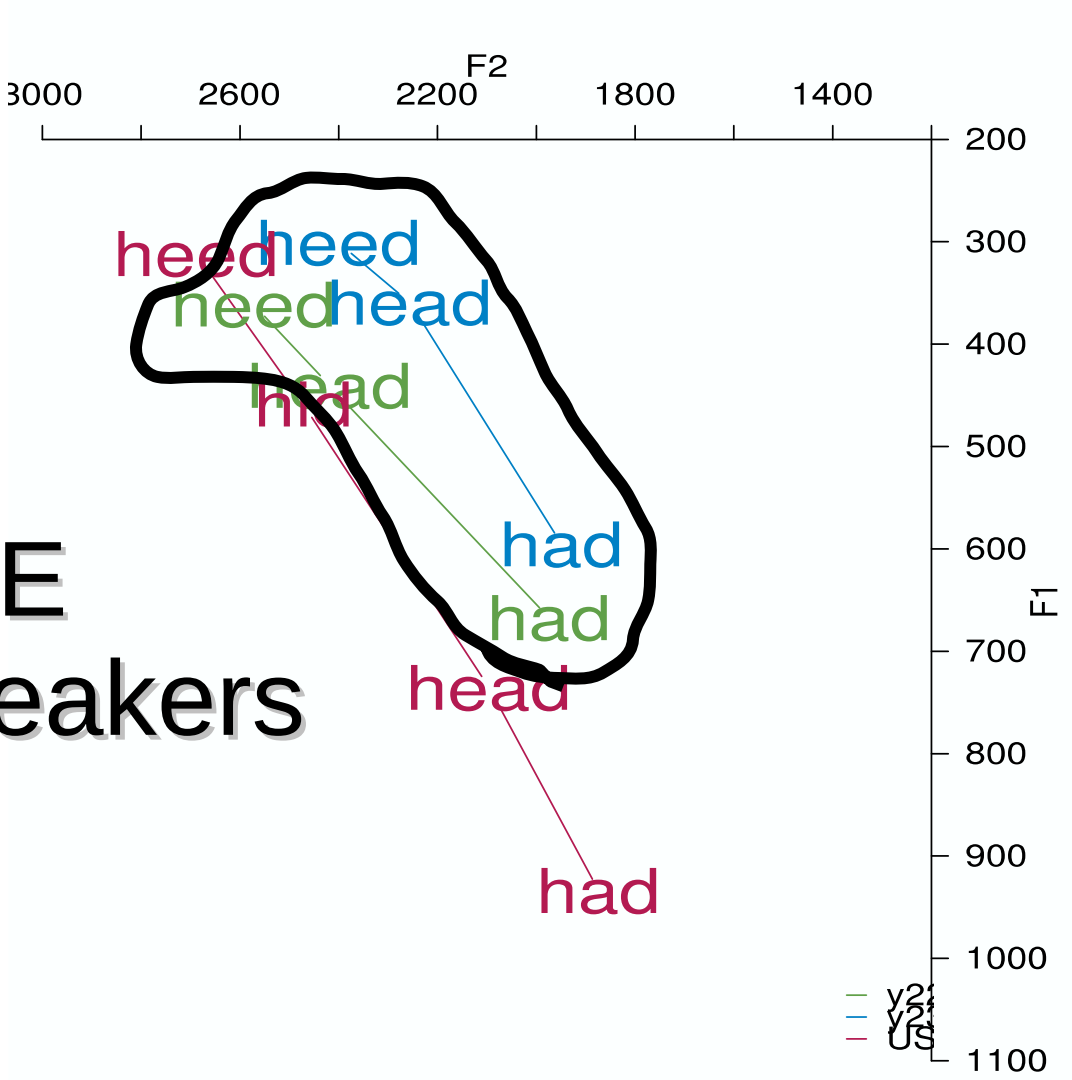
NZE vs USE Vowels



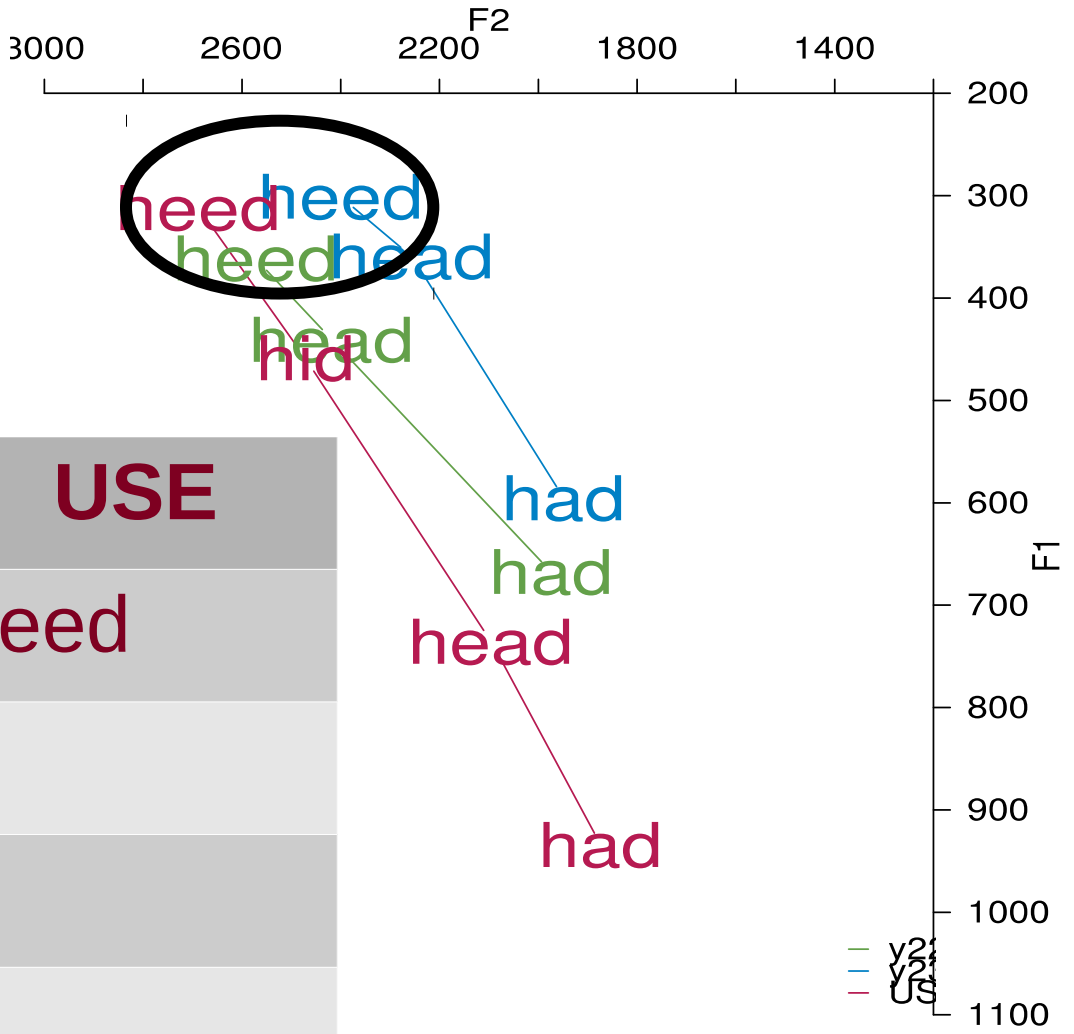
NZE vs USE Vowels



NZE vs USE Vowels

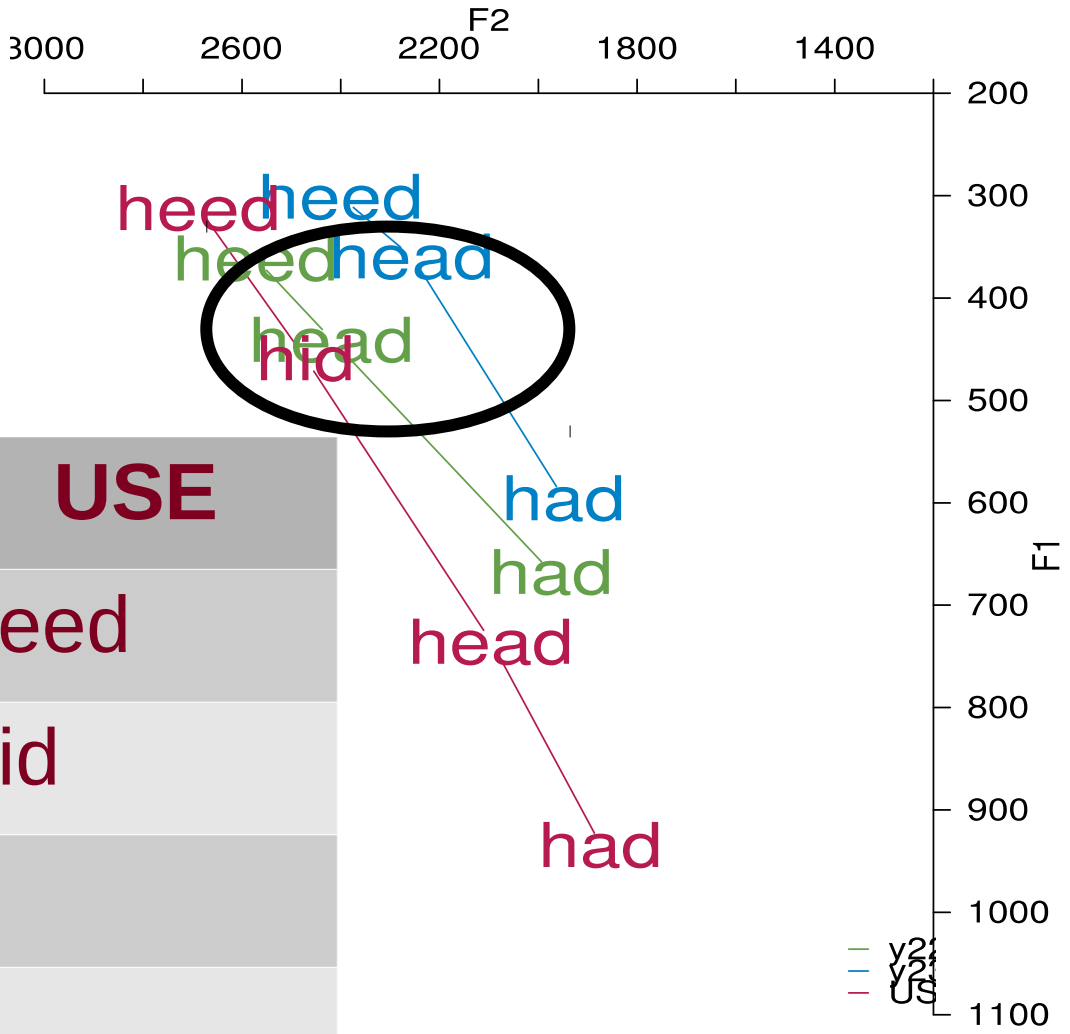


NZE vs USE Vowels



NZE	USE
heed	heed

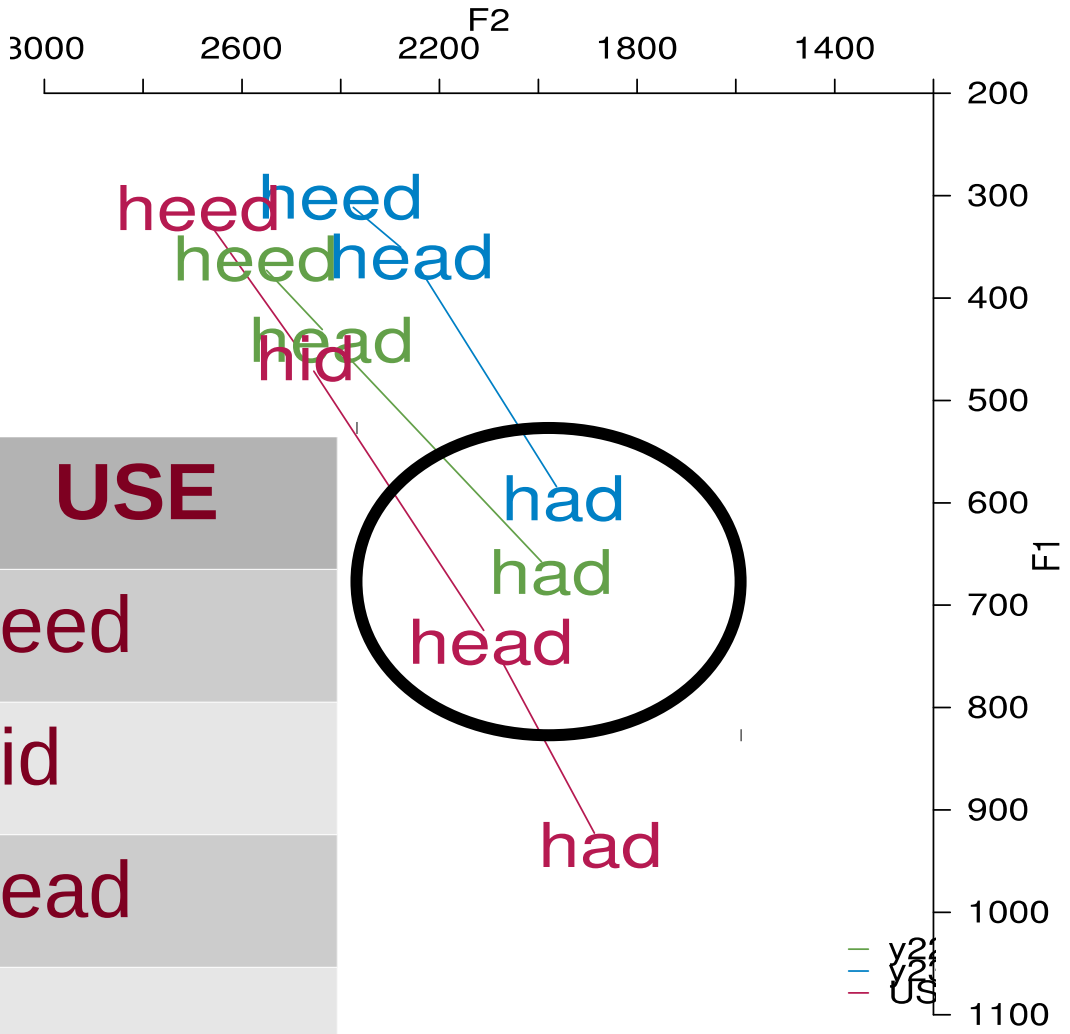
NZE vs USE Vowels



NZE	USE
heed	heed
head	hid

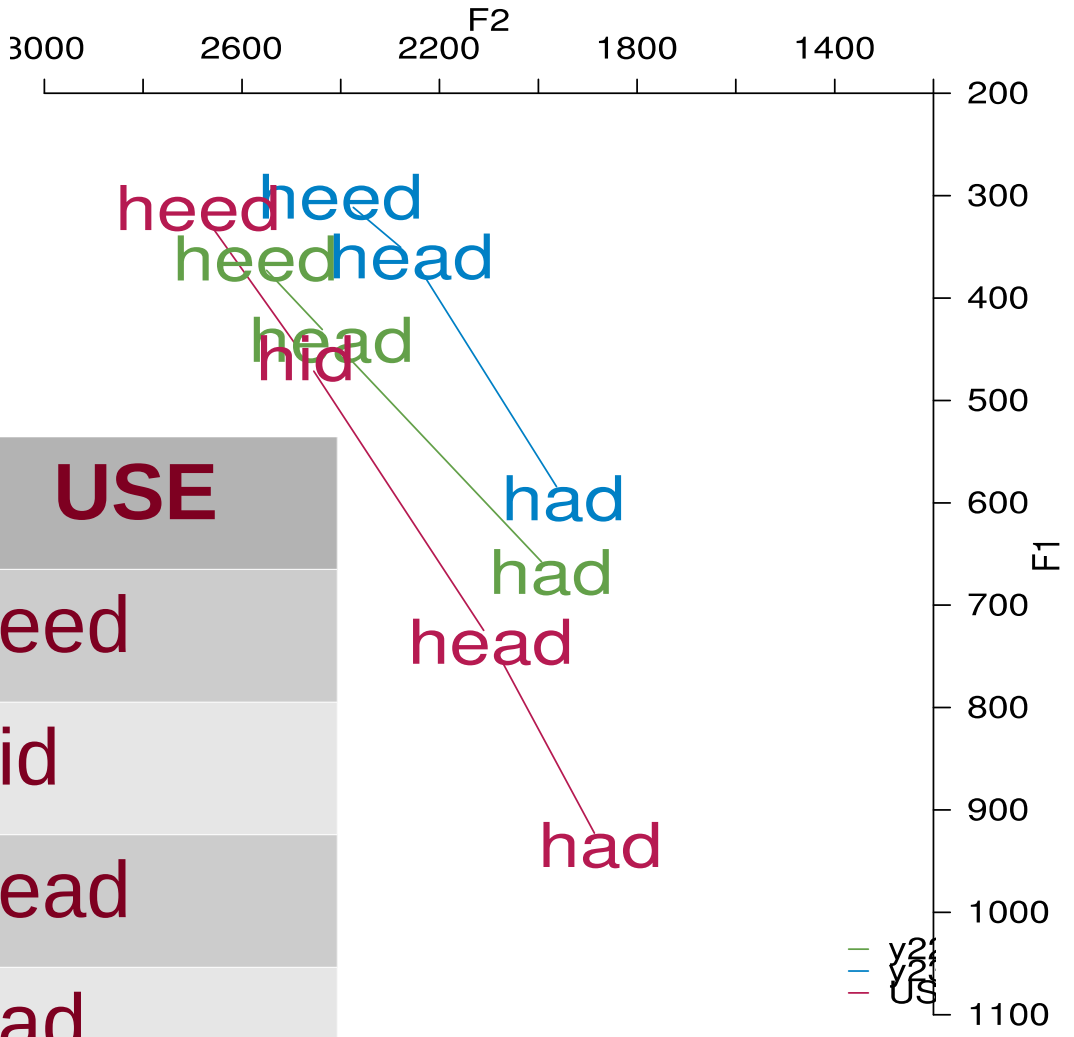
NZE vs USE Vowels

NZE	USE
heed	heed
head	hid
had	head



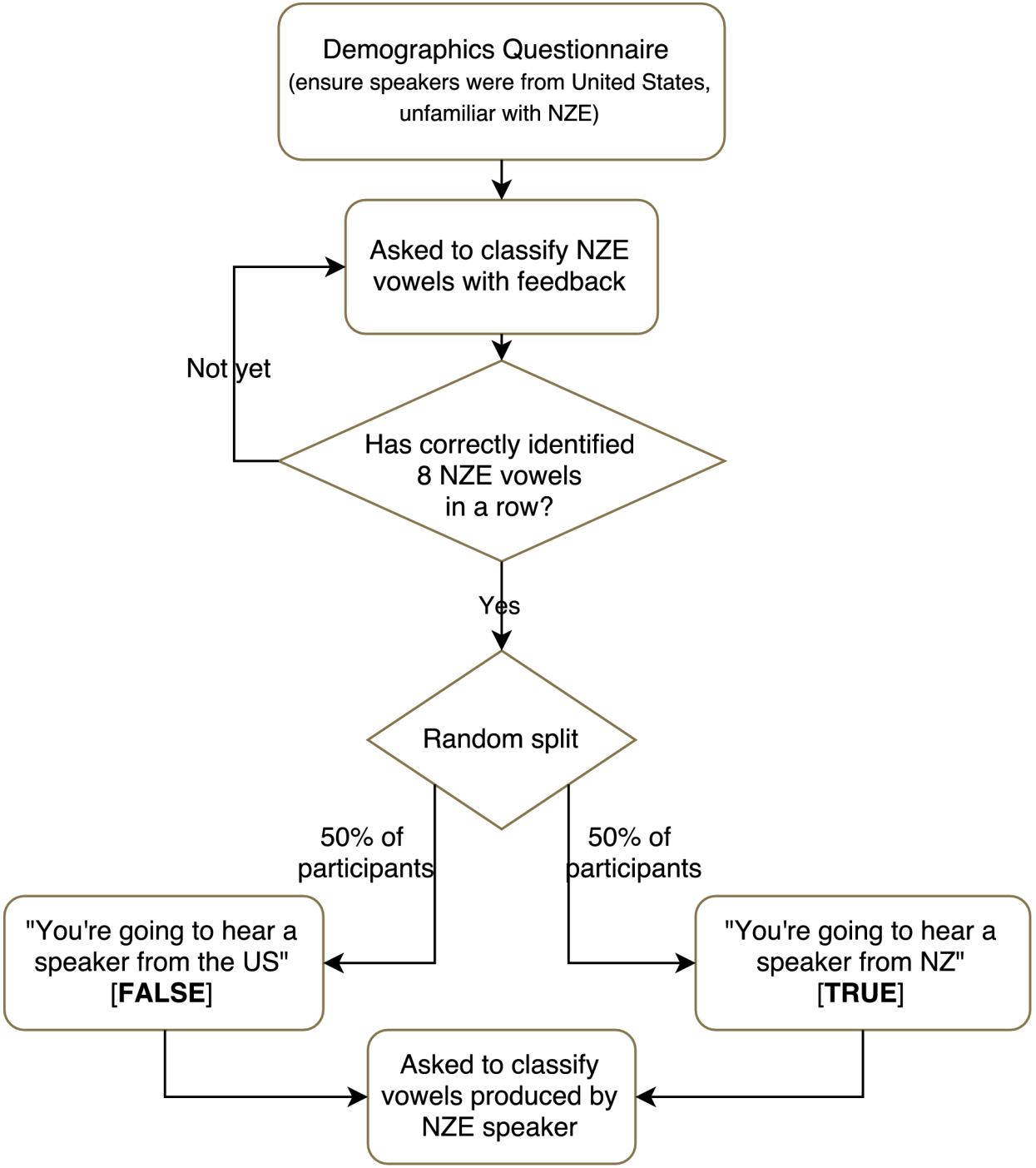
NZE and USE Vowels

NZE	USE
heed	heed
head	hid
had	head
--	had



Methodology (Specifics)

- Items:
  - “head”, “had” and “heed” tokens produced by two sociolinguistically matched native NZE speakers
    - Tokens provided by Catherine Watson (thanks!)
  - 150 ms vowel tokens taken from each
  - “hid” excluded due to duration contrast (Watson, Maclagan & Harrington 2000)
- Participants:
  - 15 native English speakers from the US who had never been to NZ
  - Experiment completed on-line using Psytoolkit (Stoet 2010)
    - Avoided interviewer effects (e.g. Hay, Warren & Drager 2006)
    - Code for experiment available on Github


Demographics Questionnaire
(ensure speakers were from United States, unfamiliar with NZE)
Asked to classify NZE vowels with feedback
Not yet
Has correctly identified 8 NZE vowels in a row?
Yes
Random split
50% of participants
50% of participants
"You're going to hear a speaker from the US" [FALSE]
"You're going to hear a speaker from NZ" [TRUE]
Asked to classify vowels produced by NZE speaker

Results 1:
Training Portion

Training Data

(All participants together)

	had	head	heed	hid
had	156	10	1	0
head	26	127	29	0
heed	3	35	147	0
hid	7	32	20	0

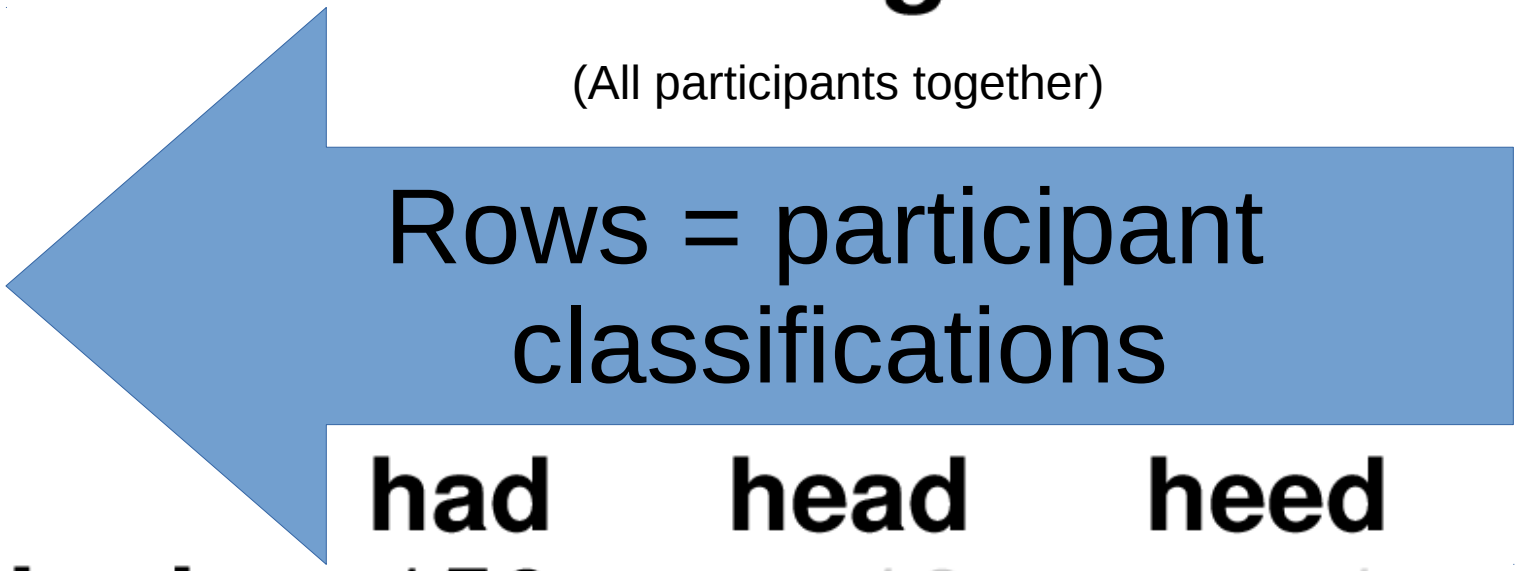
Training Data

Columns = correct classifications

	had	head	heed	hid
had	156	10	1	0
head	26	127	29	0
heed	3	35	147	0
hid	7	32	20	0

Training Data

(All participants together)



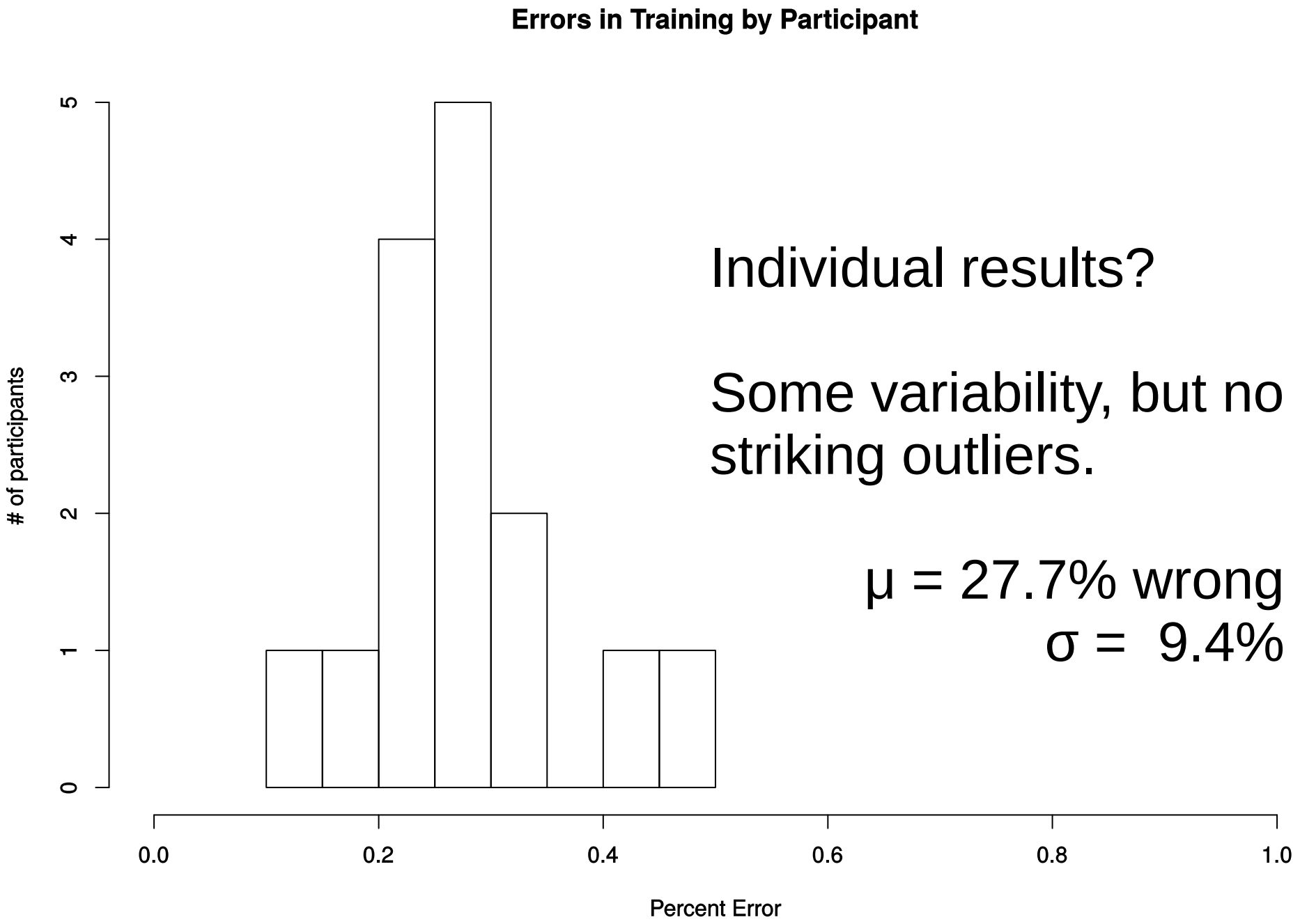
	had	head	heed	hid
had	156	10	1	0
head	26	127	29	0
heed	3	35	147	0
hid	7	32	20	0

Training Data

(All participants together)

Most classifications correct
(Cohen's Kappa = 0.61)

	had	head	heed	hid
had	156	10	1	0
head	26	127	29	0
heed	3	35	147	0
hid	7	32	20	0



Individual results?

Some variability, but no striking outliers.

μ = 27.7% wrong
σ = 9.4%

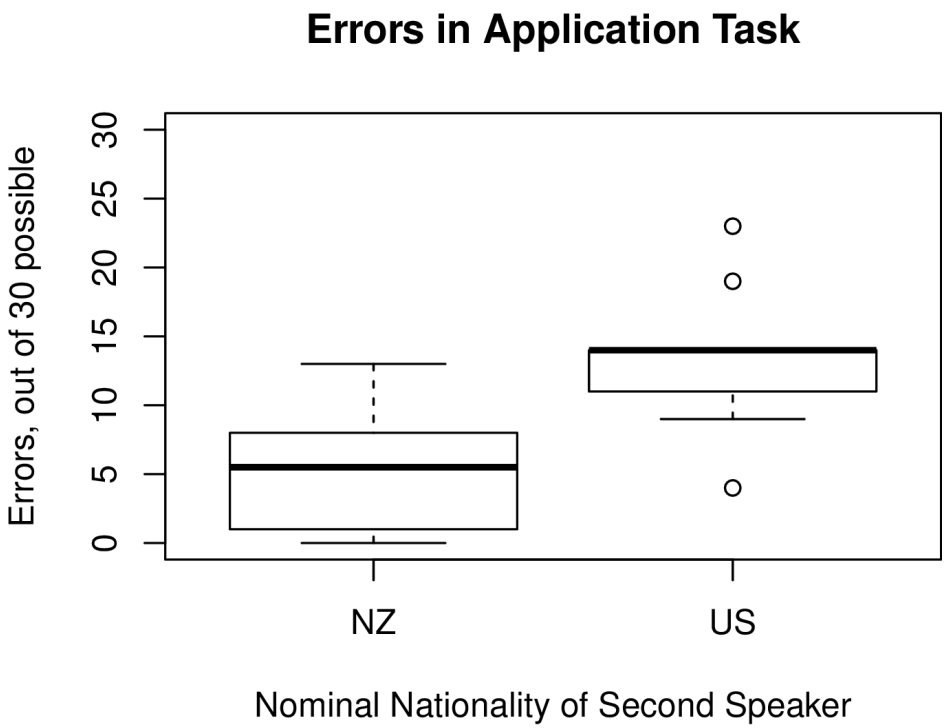
Main takeaway:

Pretty much everyone learned to correctly identify the NZE English vowels.

If they fail to do it during the testing portion we know it's not because they *can't*.

Results 2:
Testing Portion
(With Social Information)

Were the Groups Different?



- Yes!
  - NZ group much more accurate (K = 0.66)
  - US group made more errors (K = 0.48)
  - Significant difference ($\chi^2(477) = 14.6$, $p < 0.01$)
- Social information is definitely changing classifications
- But how?

Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	16	0
heed	3	16	57	0
hid	0	13	6	0

NZ group has pretty much the same classifications as in training

Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

US group has similar classifications for heed

Classifications by Each Group

US Group

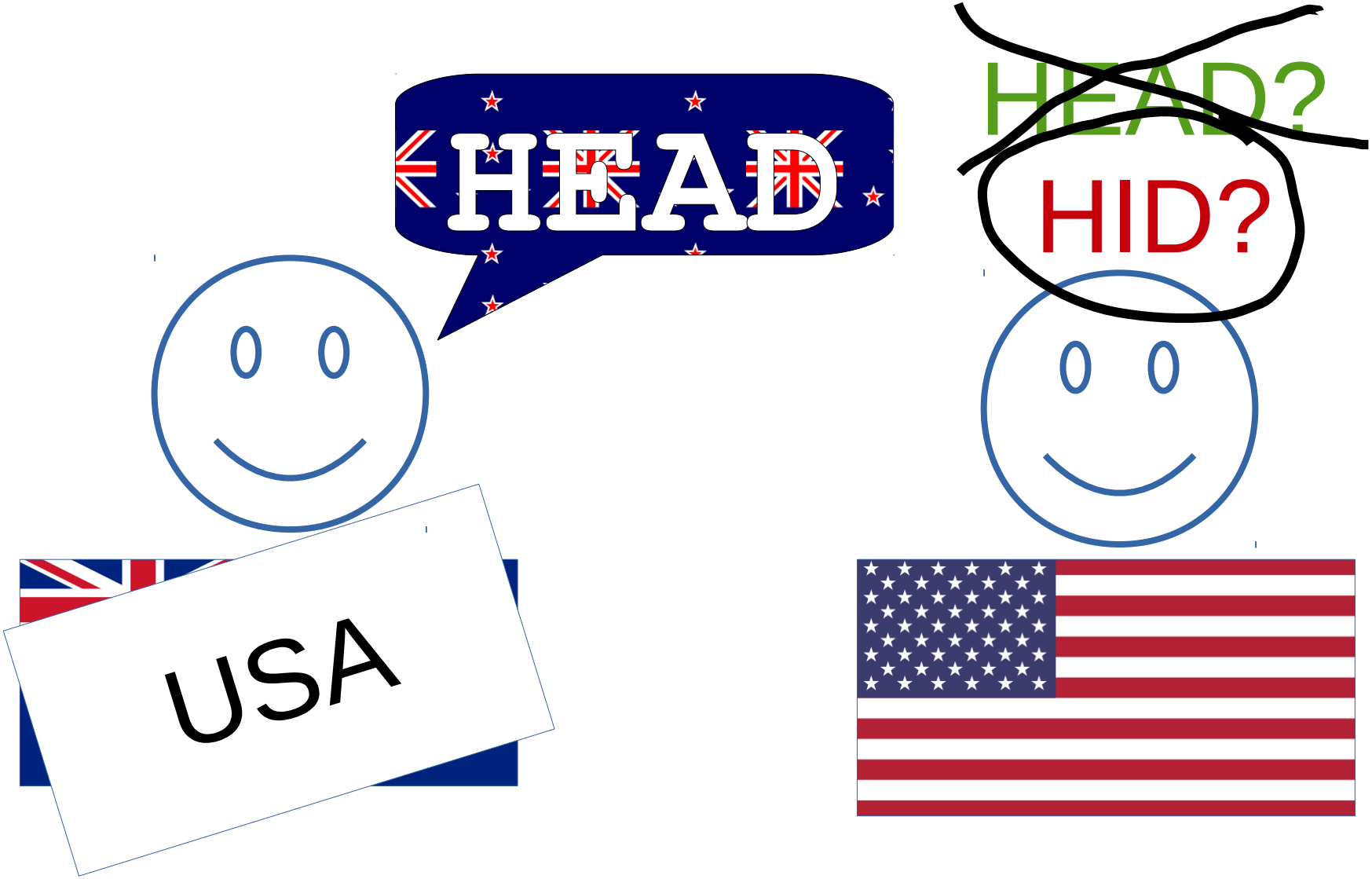
	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

NZE “head” overwhelmingly classified as USE “hid”

Methodology (Overview)



Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

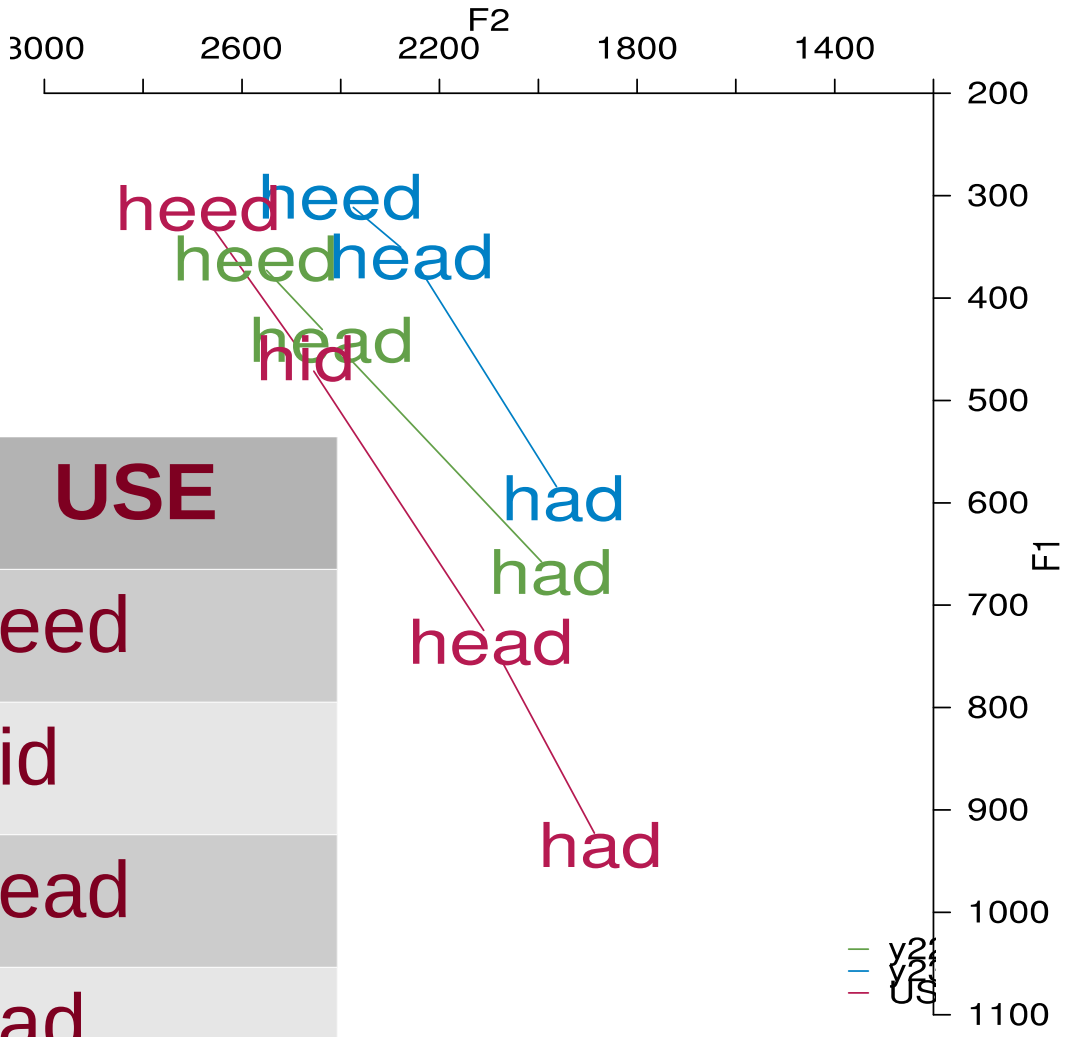
NZ Group

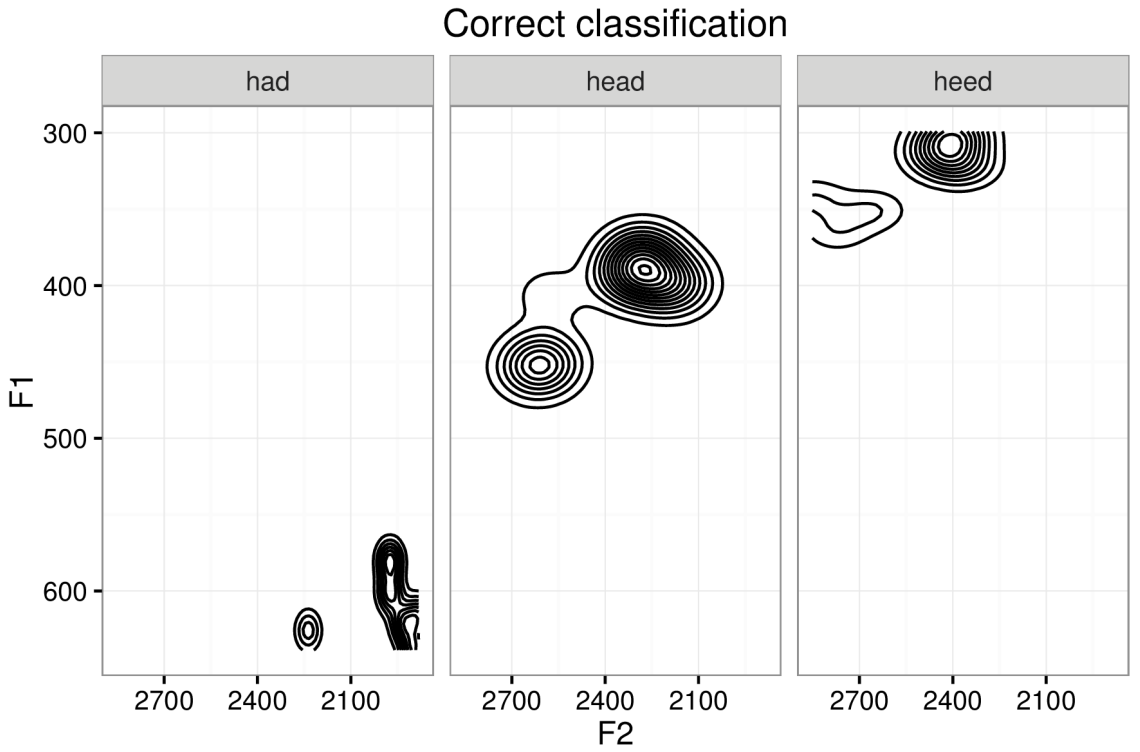
	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

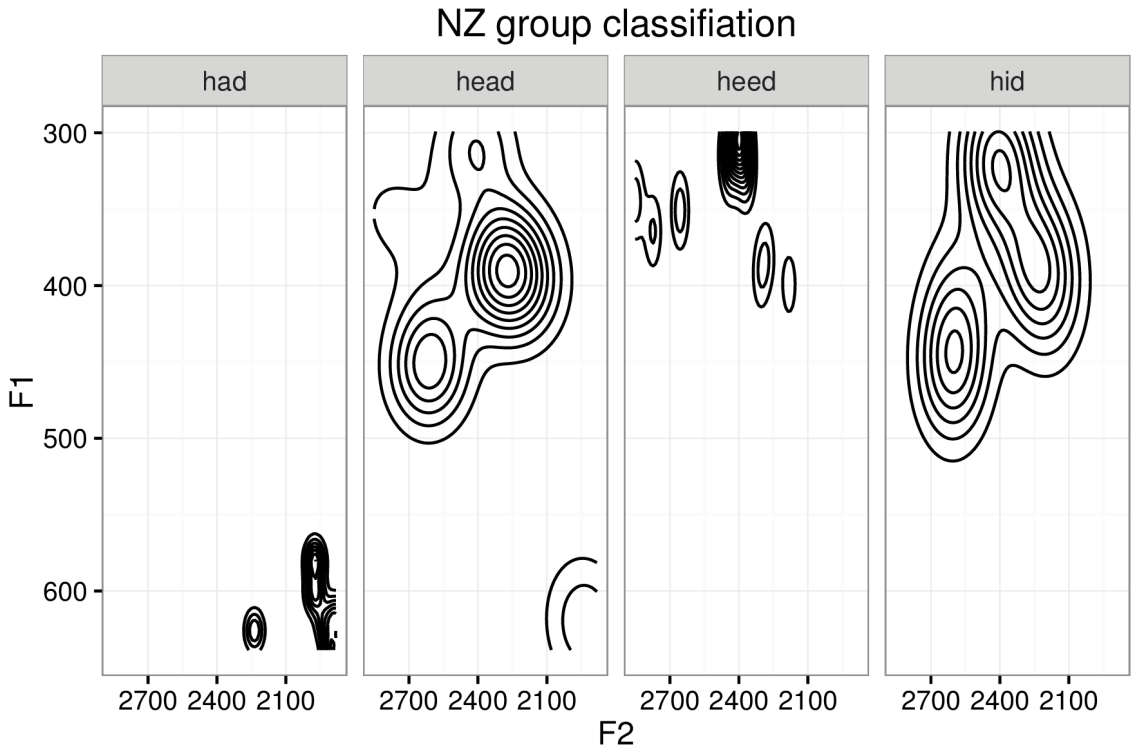
"had" slightly more likely to be classified as "head"

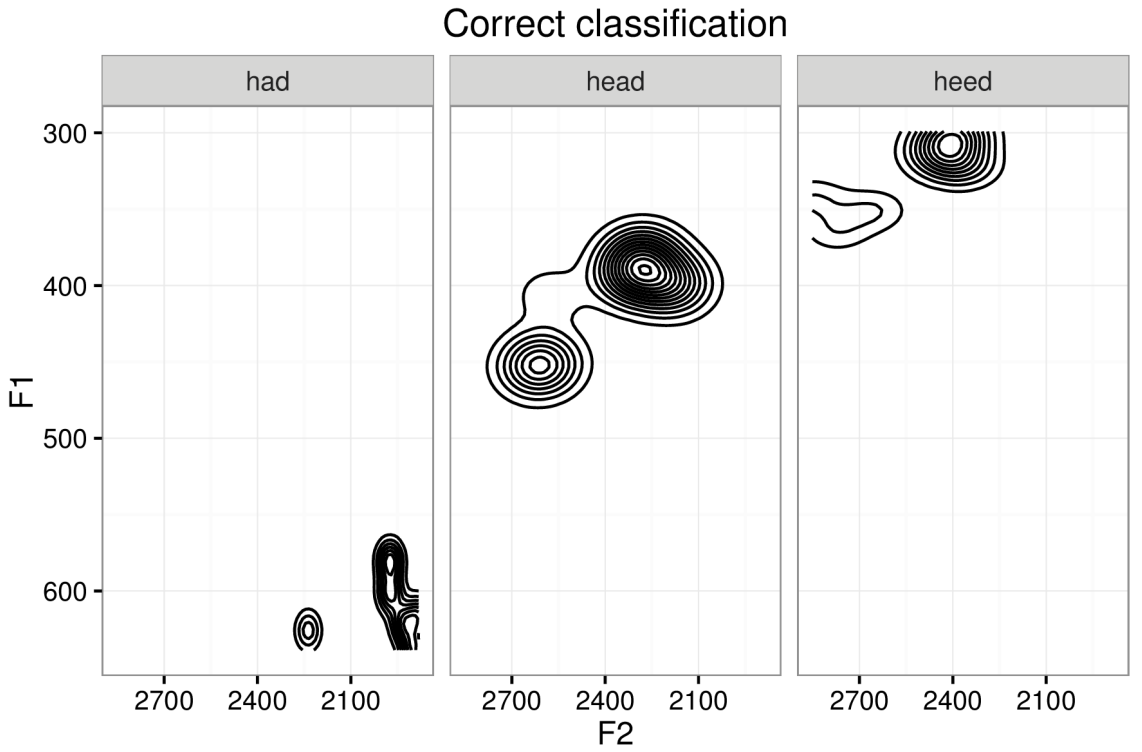
NZE and USE Vowels

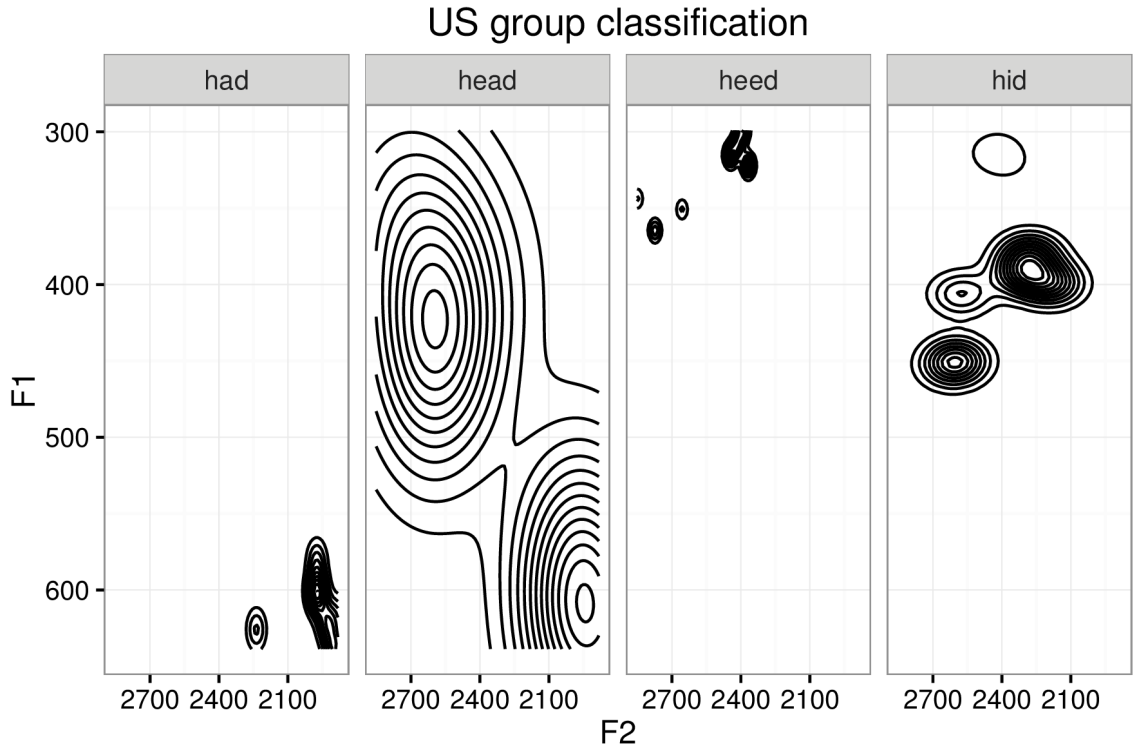
NZE	USE
heed	heed
head	hid
had	head
--	had




Correct classification
had
head
heed
300
400
500
600
F1
2700
2400
2100
F2


NZ group classifiation
had
head
heed
hid
300
400
500
600
F1
2700 2400 2100
F2


Correct classification
had
head
heed
300
400
500
600
F1
2700
2400
2100
F2


US group classification
had
head
heed
hid
300
400
500
600
F1
2700 2400 2100
F2

Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

Classifications by Each Group

US Group

NZ Group

	had	head	heed	hid
had	76	0	0	0
head	8	50	10	0
heed	3	16	57	0
hid	0	13	6	0

Classifications by Each Group

US Group

	had	head	heed	hid
had	55	0	0	0
head	22	22	4	0
heed	0	4	64	0
hid	0	56	7	0

NZ Group

Classifications consistent with social information (USE)

Two possibilities:

- ~~Social information doesn't matter, listeners rely on acoustics~~
- Social information does matter, can override acoustics

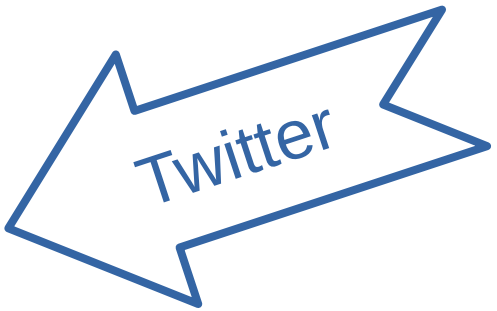
Conclusion

- Social information affects the application of perceptual learning
  - Recent perceptual learning won't be applied if social knowledge suggests it shouldn't
- Social knowledge may help explain why sometimes perceptual learning isn't carried over (Levi et al. 2011, Eisner & McQueen 2005)
- Next steps:
  - Native vs. non-native dialect
  - Change amount of acoustic information (longer samples)
  - Explore accuracy/precision of classifications

Thanks!

Questions?

Slides: http://tinyurl.com/EXAPP-Tatman
Email: rctatman@uw.edu



References

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*,*116*(6), 3647-3658.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & psychophysics*, *67*(2), 224-238.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865-892.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458-484.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262-268.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts how listeners adjust to speaker variability. *Psychological science*,*19*(4), 332-338.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible?. *The Journal of the Acoustical Society of America*, *130*(6), 4053-4062.

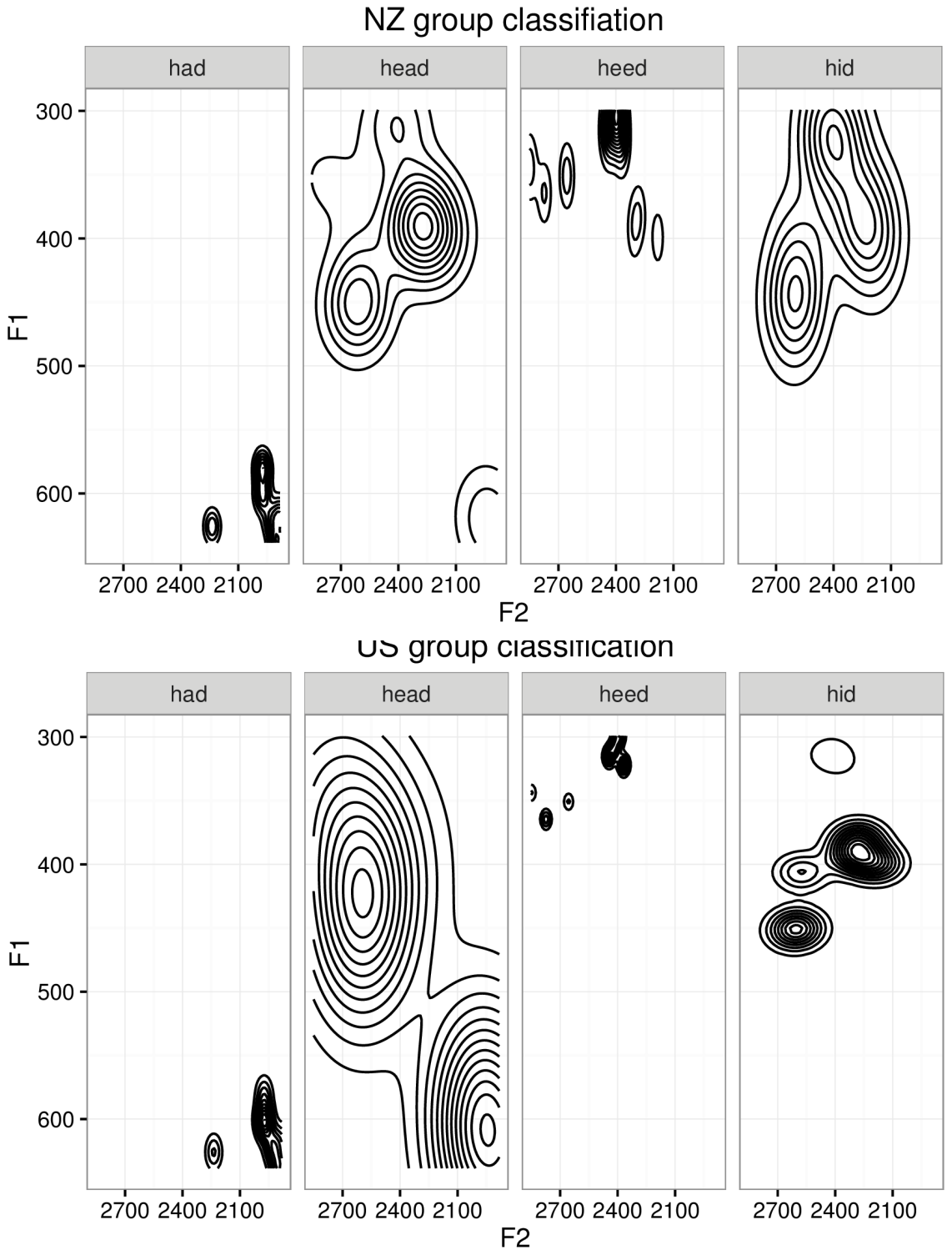
Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology*, *18*(1), 62-85.

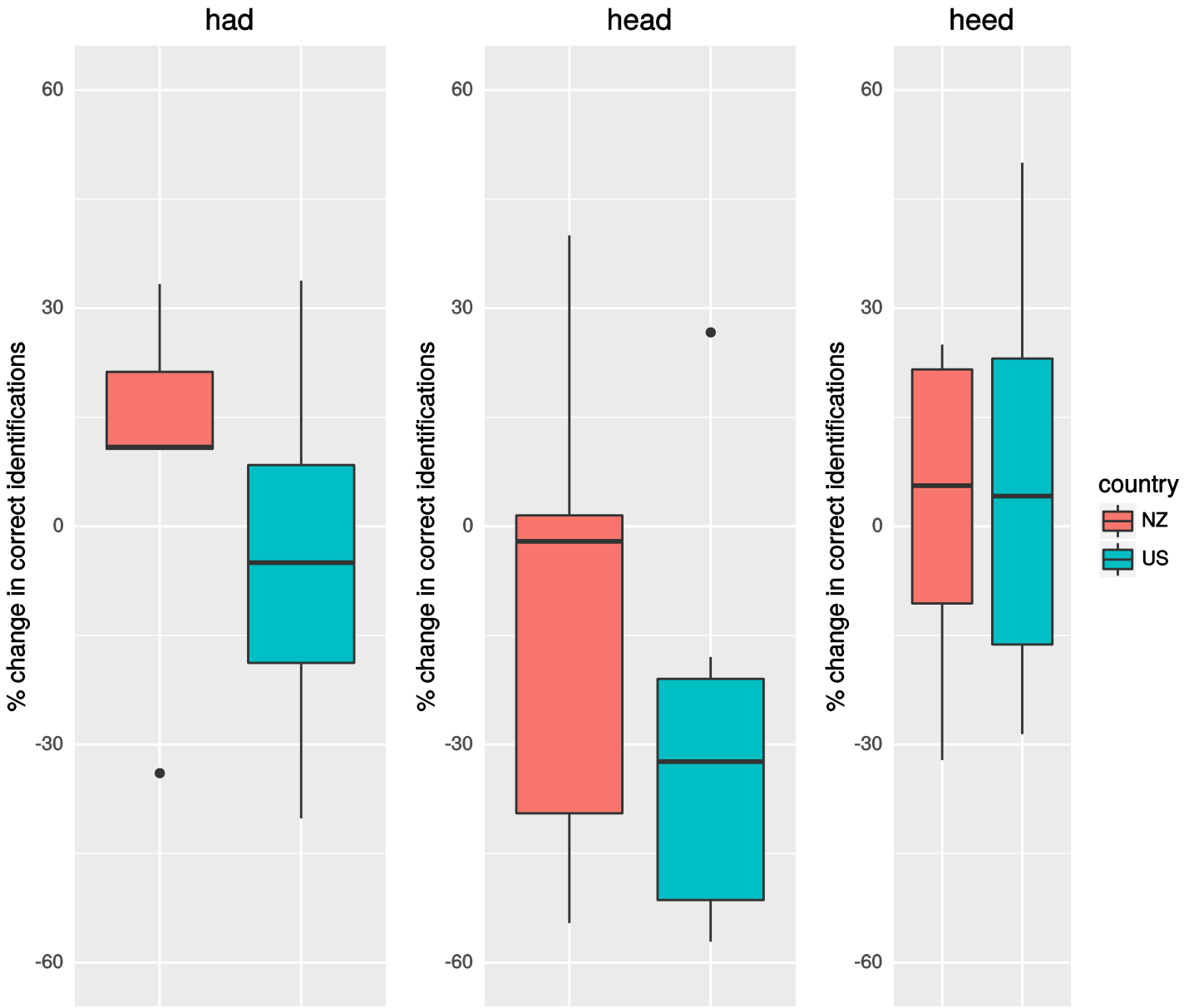
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096-1104.

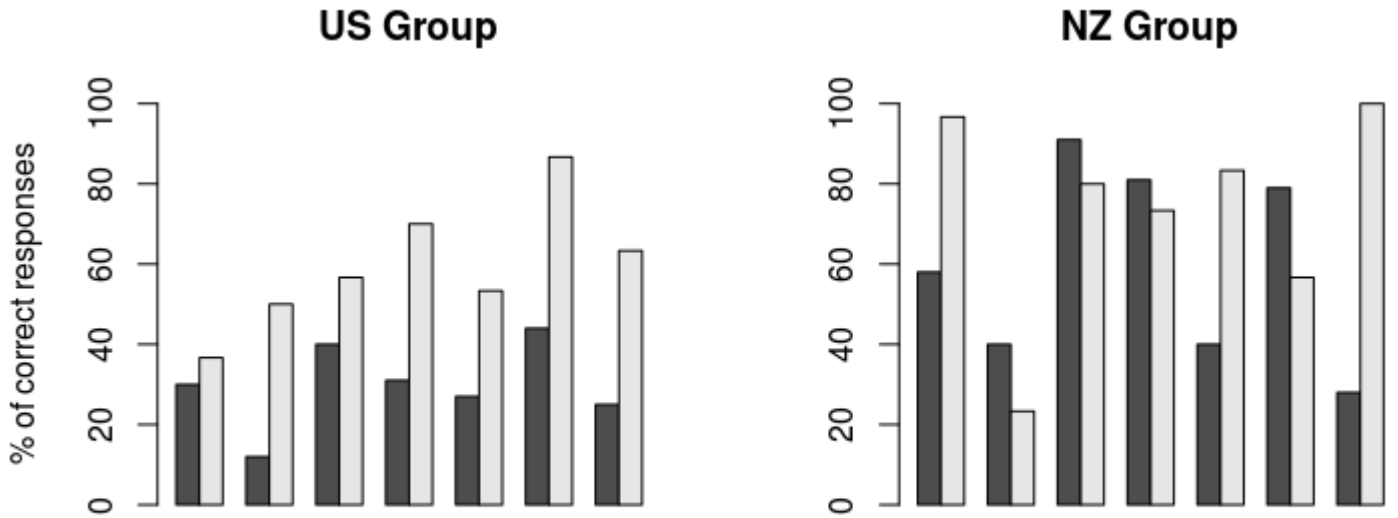
Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of language and social psychology*, *18*(1), 86-100.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language variation and change*, *12*(01), 51-68.


NZ group classifiation
had
head
heed
hid
300
400
500
600
F1
2700 2400 2100
F2
US group classification
had
head
heed
hid
300
400
500
600
F1
2700 2400 2100
F2


had
head
heed
% change in correct identifications
60
30
0
-30
-60
country
NZ
US

Individual Results



Training (dark) and testing (light), by subject and group