Clustering techniques for phonetic categories and their implications for phonology

William J. Idsardi
University of Maryland
JHU, July 16, 2012

Collaborators

- Brian Dillon, University of Massachusetts
- Ewan Dunbar, University of Maryland

Outline

- main message: there is no surface
- introduction: 2 anecdotes
- 3 eg's: speakers, contexts, locus equations
- some implications

2 anecdotes

Consilience

- between theories, models, experiments
- Geoff Hinton on layers
- U-shaped curve
- theory: phonemes, features, opacity

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Sparse Codes for Speech Predict Spectrotemporal Receptive Fields in the Inferior Colliculus

Nicole L. Carlson[1,2], Vivienne L. Ming[1], Michael Robert DeWeese[1,2,3]*

1 Redwood Center for Theoretical Neuroscience, University of California, Berkeley, California, United States of America, **2** Department of Physics, University of California, Berkeley, California, United States of America, **3** Helen Wills Neuroscience Institute, University of California, Berkeley, California, United States of America

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Excitatory Local Interneurons Enhance Tuning of Sensory Information

Collins Assisi[1], Mark Stopfer[2], Maxim Bazhenov[1]*

1 Department of Cell Biology and Neuroscience, University of California, Riverside, California, United States of America, **2** US National Institutes of Health, National Institute of Child Health and Human Development, Bethesda, Maryland, United States of America

OPEN ACCESS Freely available online PLoS one

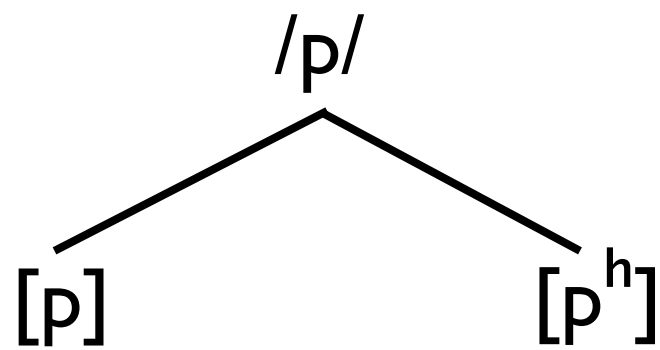
A Neural Field Model of the Somatosensory Cortex: Formation, Maintenance and Reorganization of Ordered Topographic Maps

Georgios Is. Detorakis, Nicolas P. Rougier*

INRIA CNRS: UMR 7503 Université Henri Poincaré - Nancy I Université Nancy II Institut National Polytechnique de Lorraine, Nancy, France

Theory

- Phonemes
- Allophones
- (+ features)



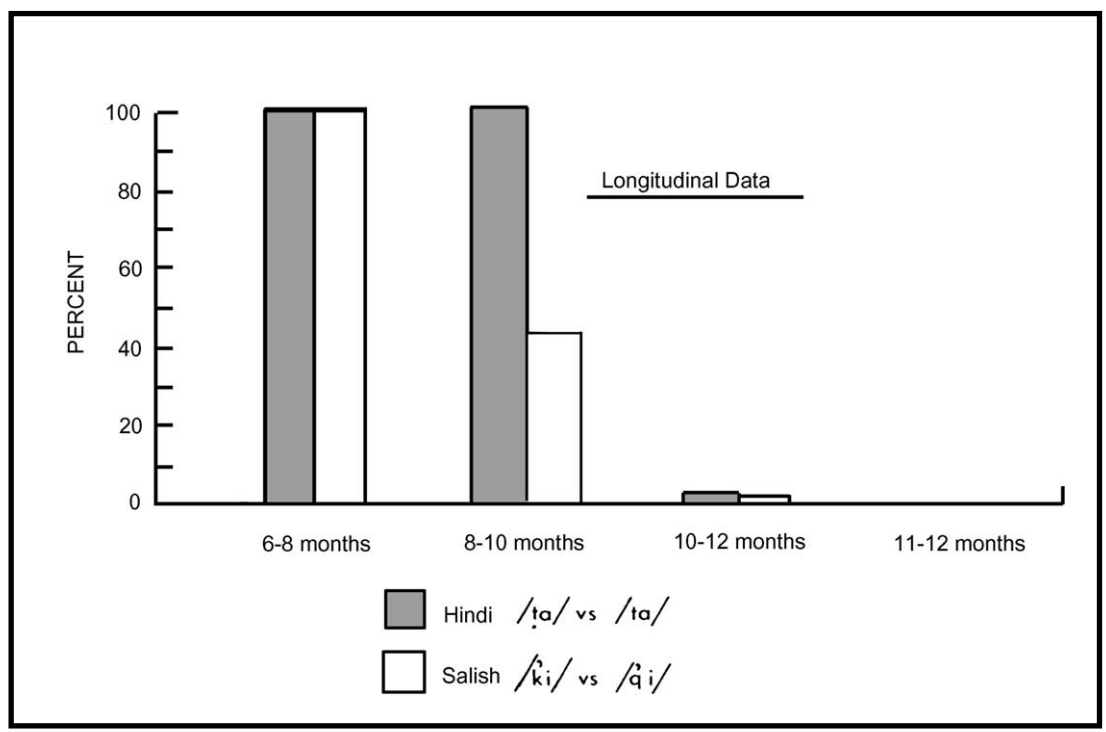
Harris, etc.

- discover phonemes through complementary distribution (discovery procedure)
- identify phones, then group into phonemes

Phones

How do experimental subjects learn phones?

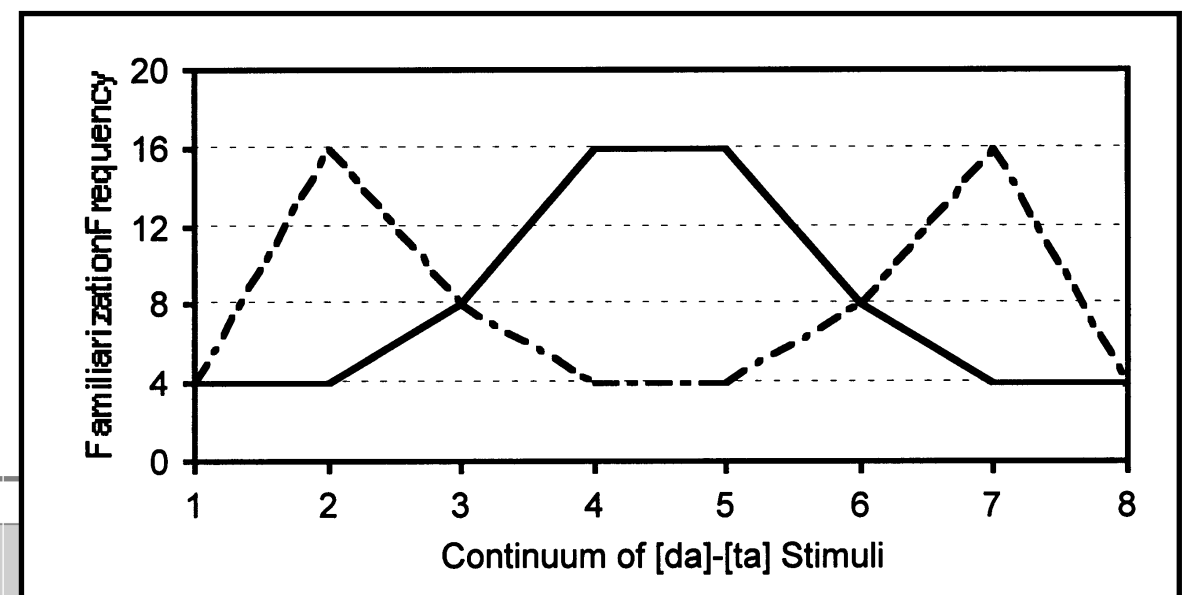
(Werker and Tees 1984)



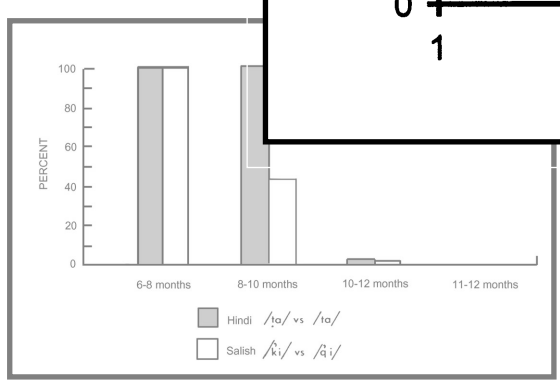
Phones

How do experimental subjects learn phones?

(Maye, Werker, and Gerken 2002)

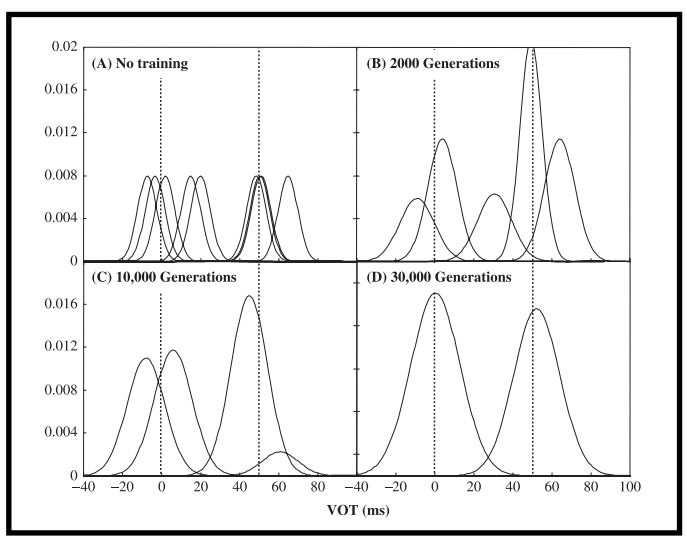


(Werker and Tees 1984)



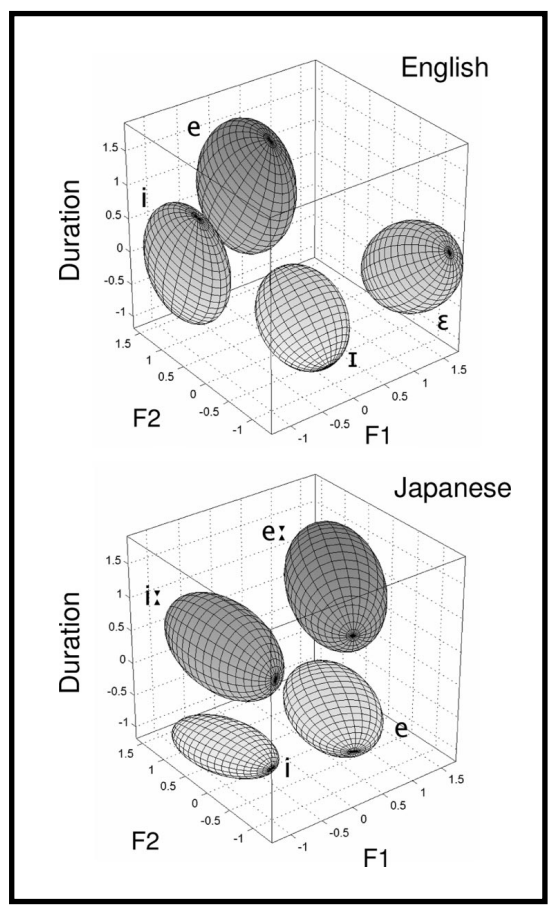
Previous statistical models

Gaussian mixture models

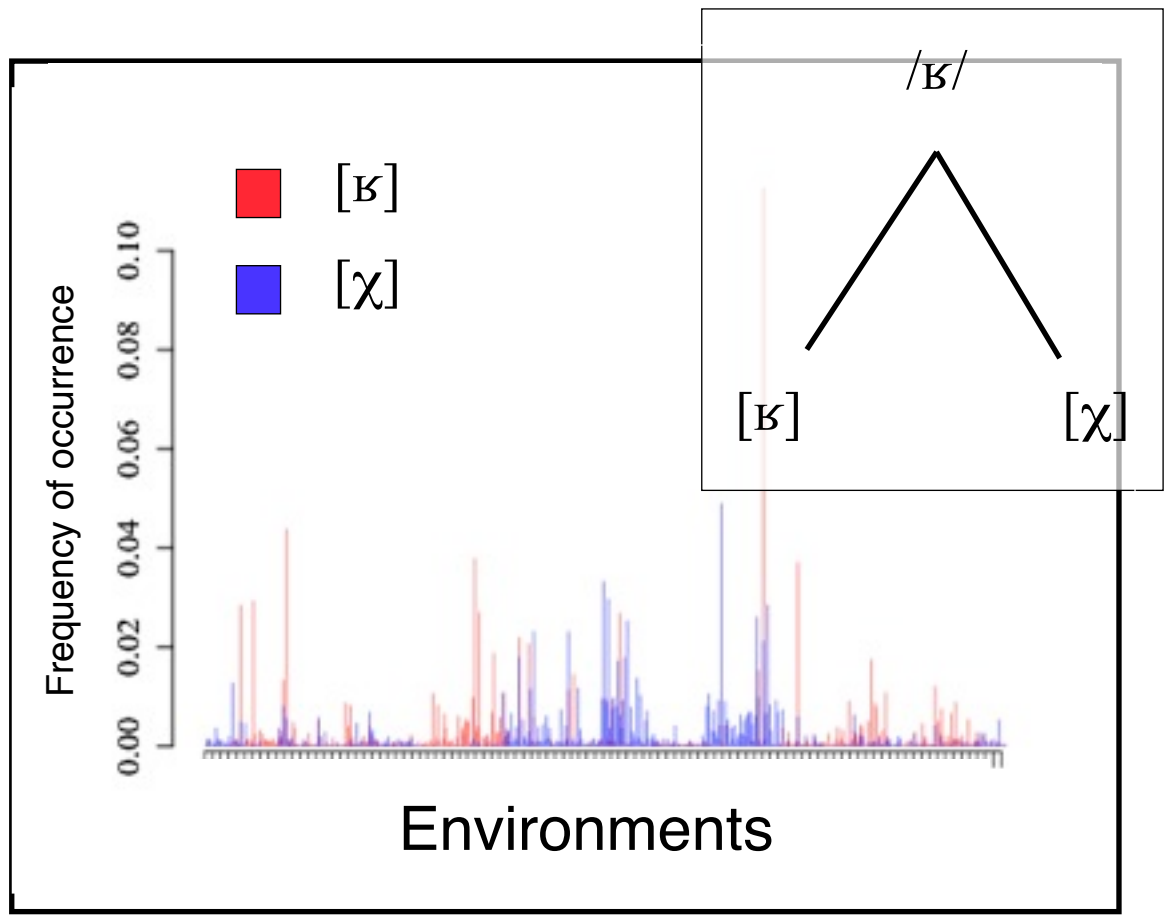


(McMurray et al. 2009)

(Vallabha et al. 2007)

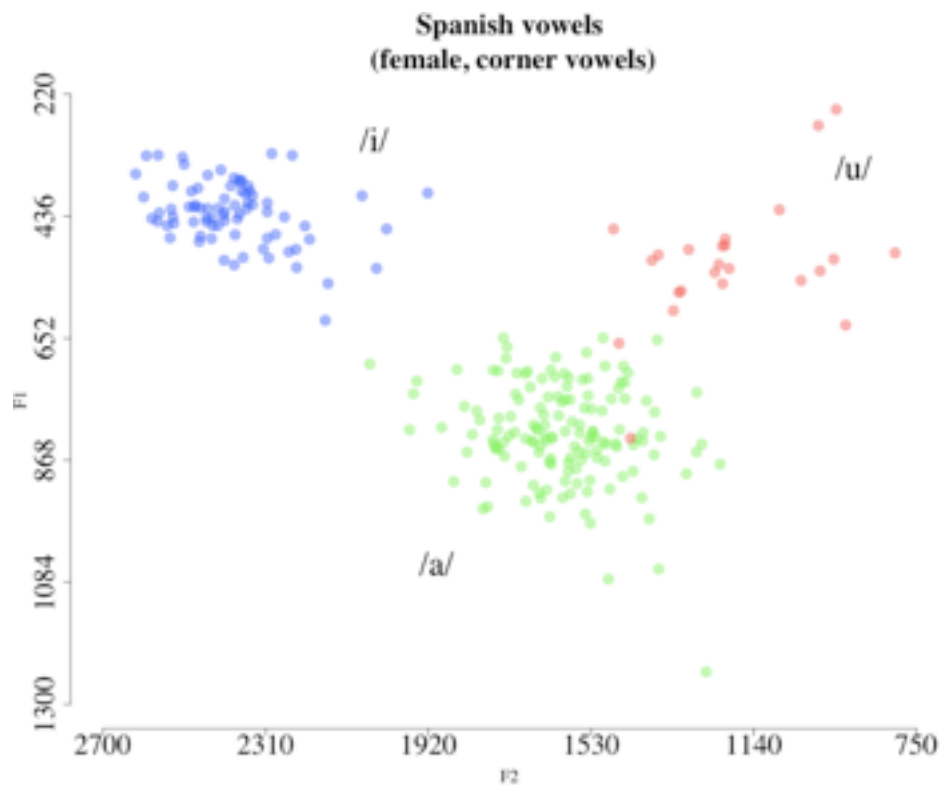


Phones to phonemes

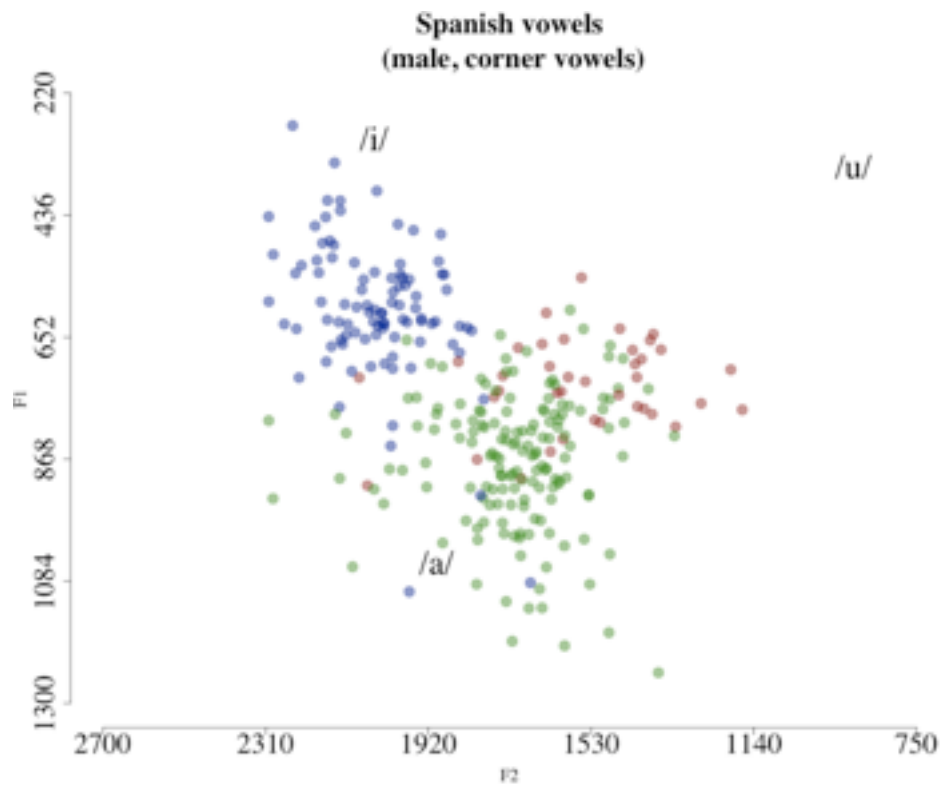


(Peperkamp et al. 2006)

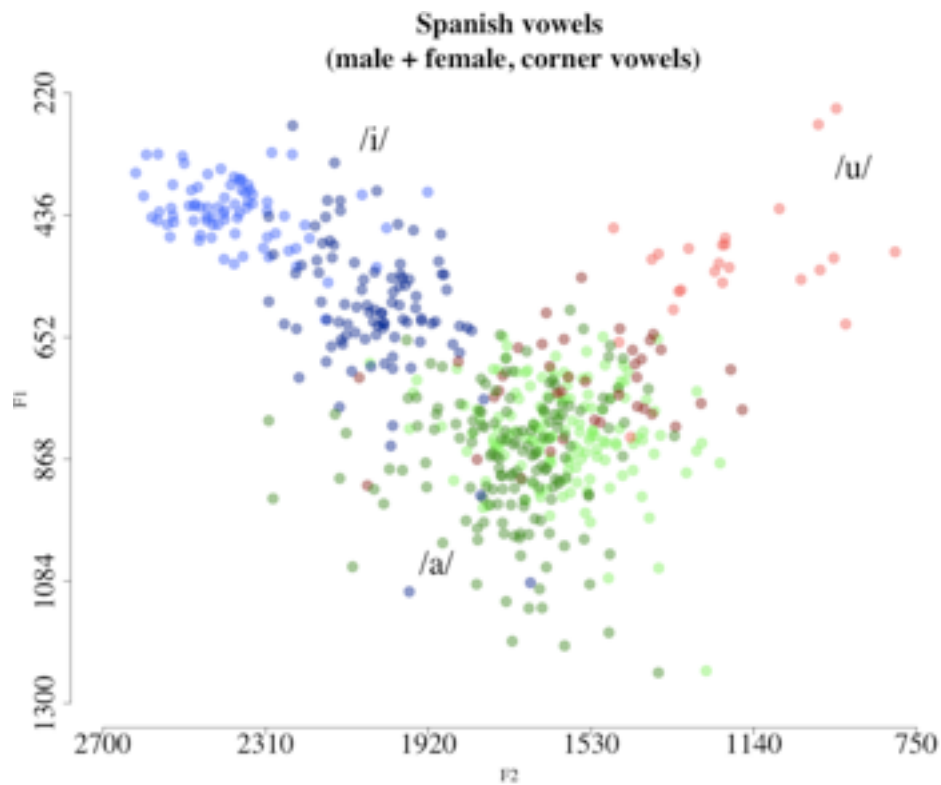
Problems


Spanish vowels
(female, corner vowels)
/i/
/u/
/a/
F1
220
436
652
868
1084
1300
2700
2310
1920
1530
1140
750
F2

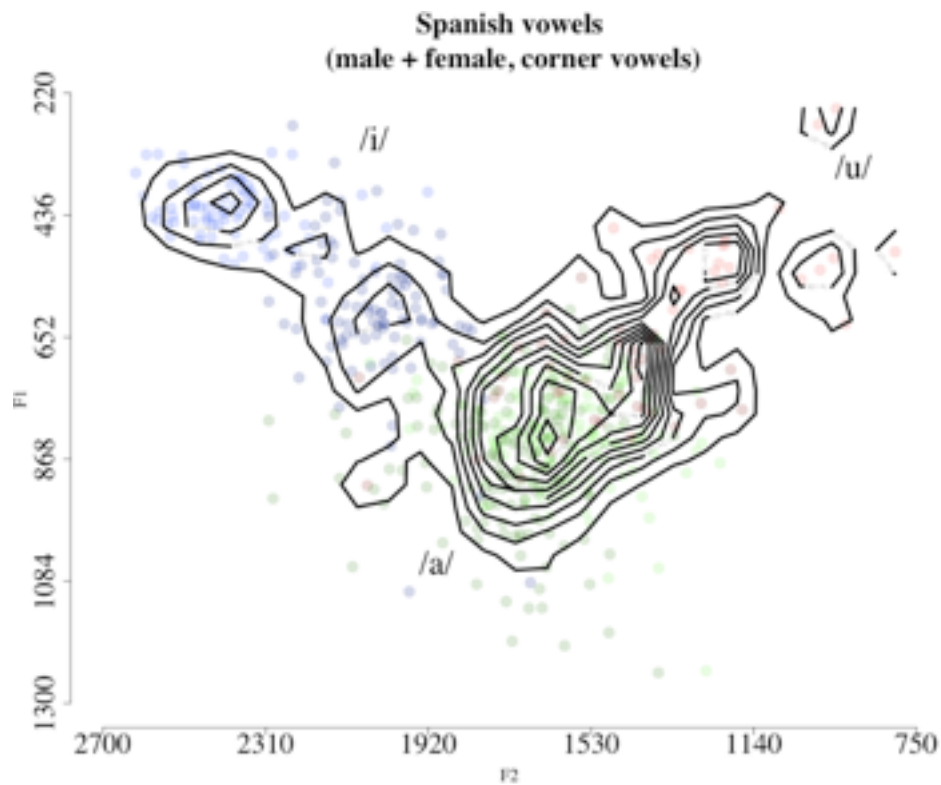
Problems


Spanish vowels
(male, corner vowels)
/i/
/u/
/a/
F1
220
436
652
868
1084
1300
2700
2310
1920
1530
1140
750
F2

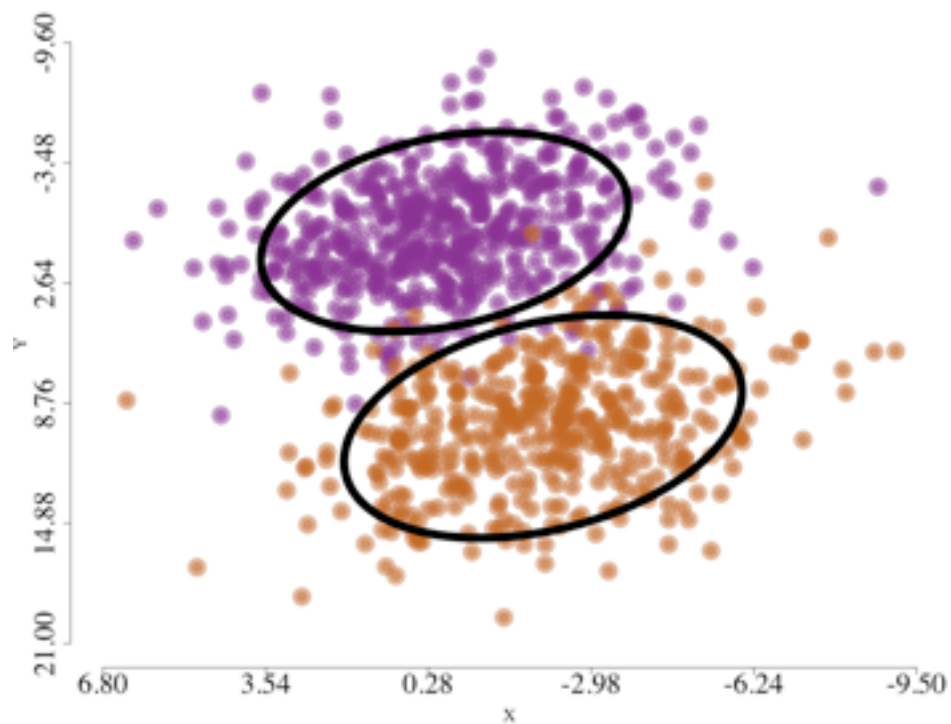
Problems


Spanish vowels
(male + female, corner vowels)
/i/
/u/
/a/
F1
220
436
652
868
1084
1300
2700
2310
1920
1530
1140
750
F2

Problems

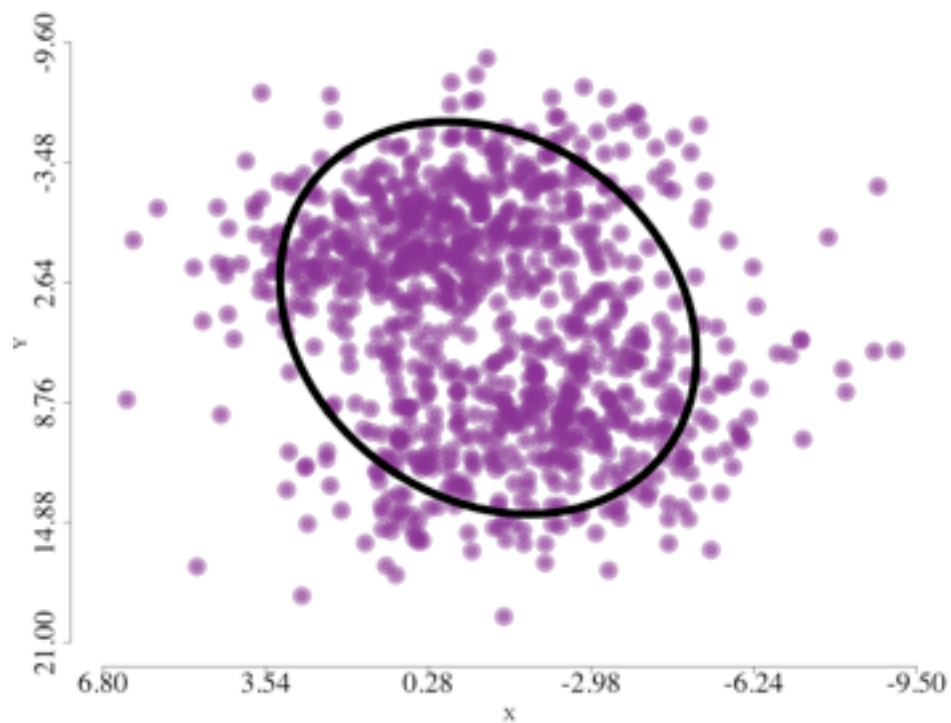

Spanish vowels
(male + female, corner vowels)
/i/
/u/
/a/
F1
220
436
652
868
1084
1300
2700
2310
1920
1530
1140
750
F2

A new model



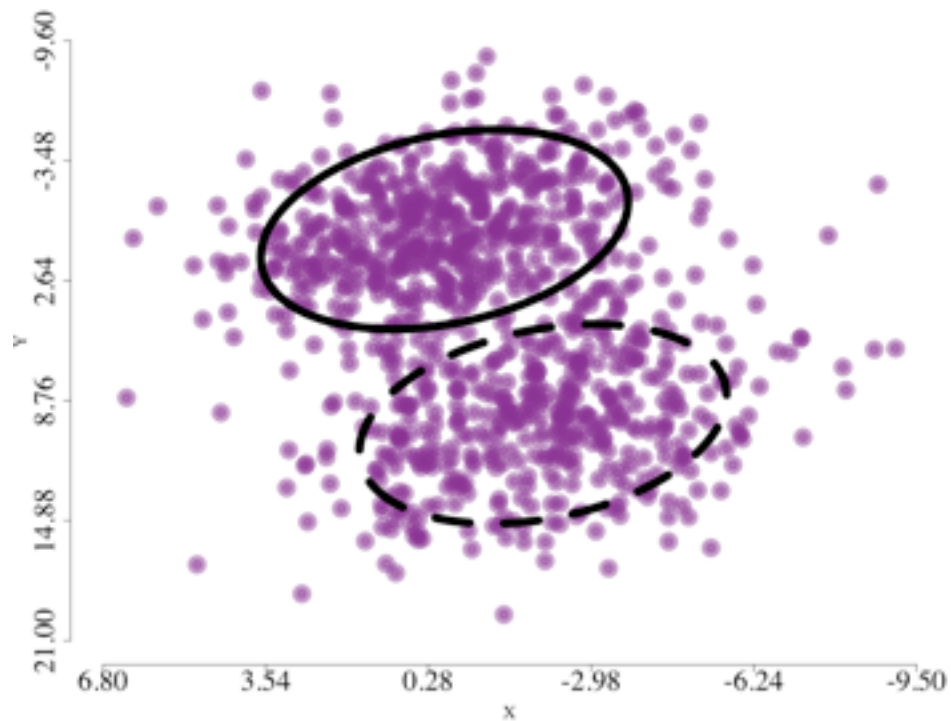
categories and transformations: c's and t's

A new model



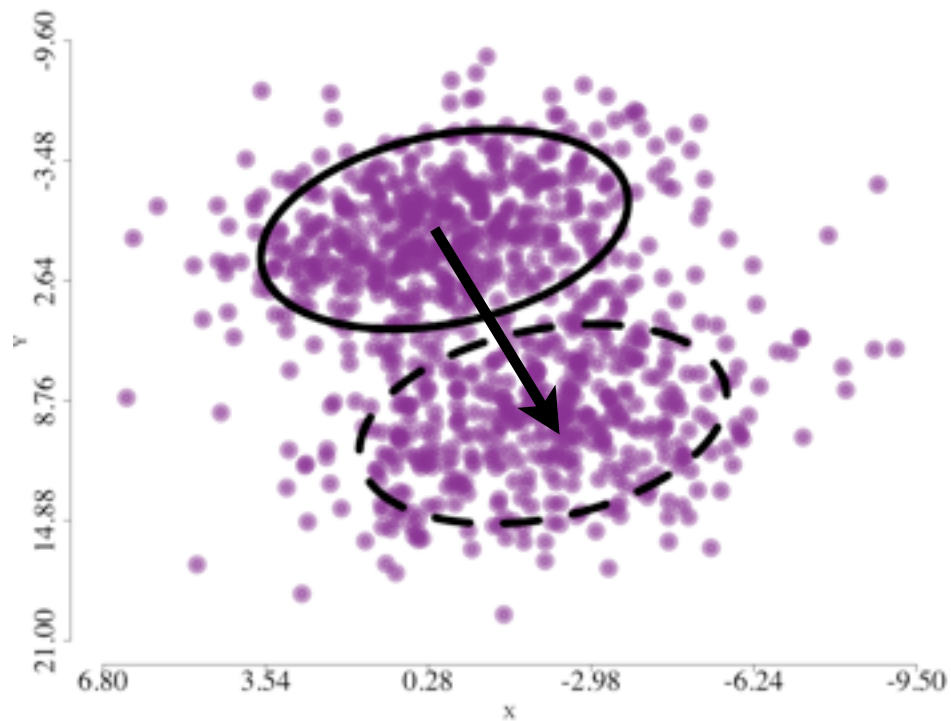
categories and transformations: c's and t's

A new model



categories and transformations: c's and t's

A new model



categories and transformations: c's and t's

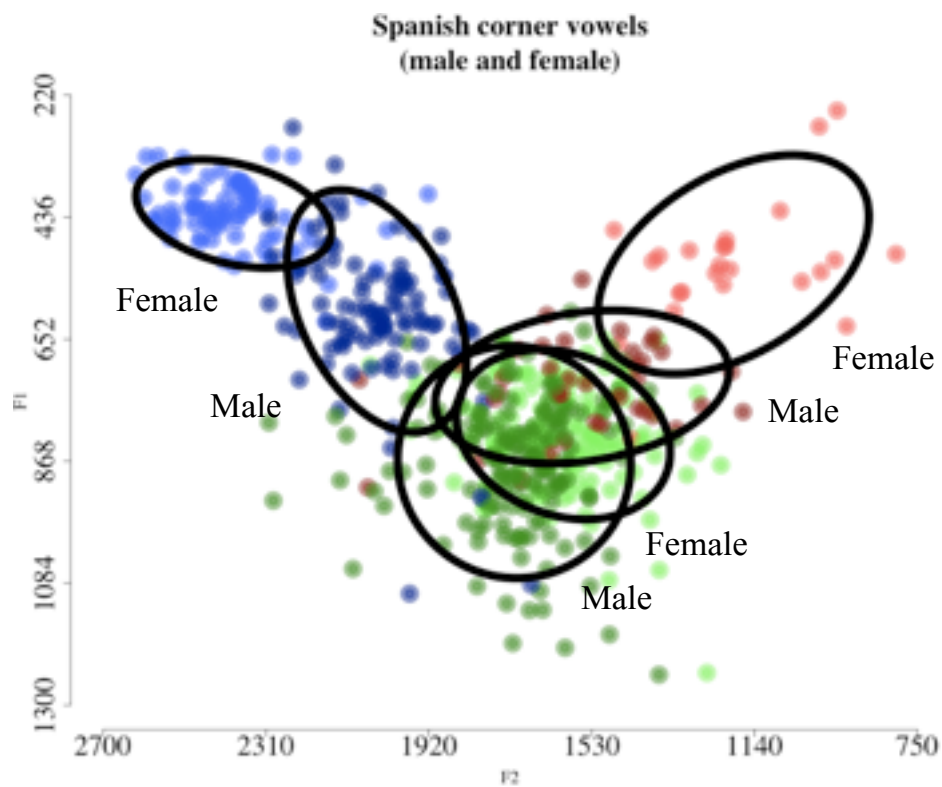
3 examples

speaker/gender

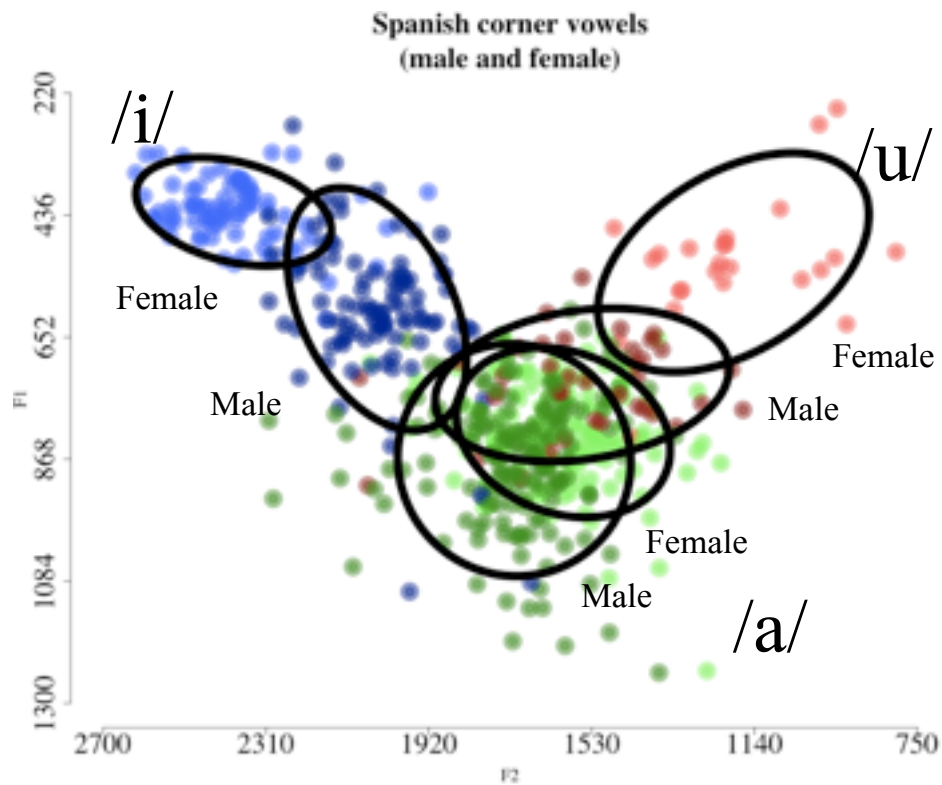
Spanish Vowels

- Data from one male and one female speaker from North America
- Vowels extracted automatically from CALLHOME telephone speech corpus using Praat: first three formants (Boersma and Weenink 2007)
- Corner vowels (/i/, /u/, /a/) extracted as test case
- 536 data points (249 female, 237 male)

Materials


Spanish corner vowels
(male and female)
F1
220
436
652
868
1084
1300
F2
2700
2310
1920
1530
1140
750
Female
Male
Female
Male
Female
Male

Materials


Spanish corner vowels
(male and female)
/i/
/u/
/a/
Female
Male
Female
Male
Female
Male
F1
220
436
652
868
1084
1300
2700
2310
1920
1530
1140
750
F2

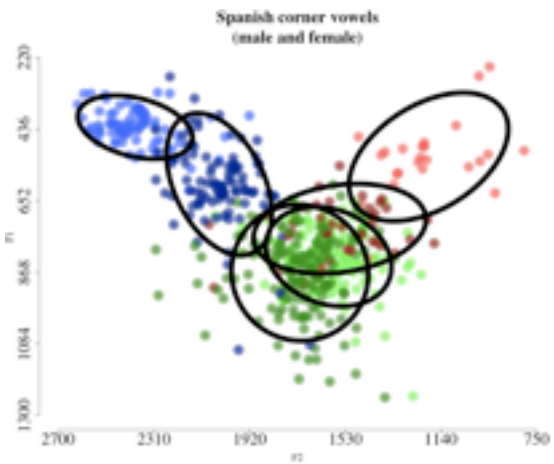
“i-Phones”

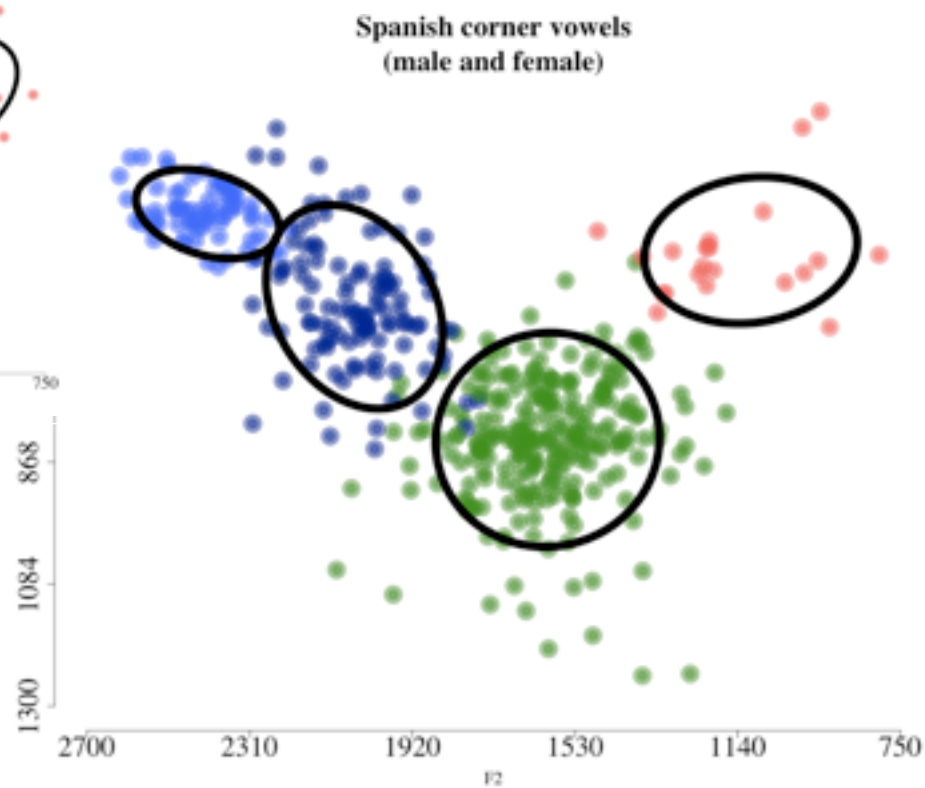
Methods

- Nonparametric Bayesian Gaussian mixture (Dirichlet process mixture); as many/few categories as the data demands
- 10-fold cross validation
- Fit 10 times, each time holding out 10% of the data for testing on new points
- Fit using MCMC (Gibbs sampler)

“i-Phones”

Results





“neither fish nor fowl”

Summary

- Speaker variability can make one lexical category look like two phonetic categories
- Speaker variability can make categories overlap
- Hard to learn any appropriate categories if you don't know about speaker variability
- We definitely want the lexical inventory to abstract out male/female

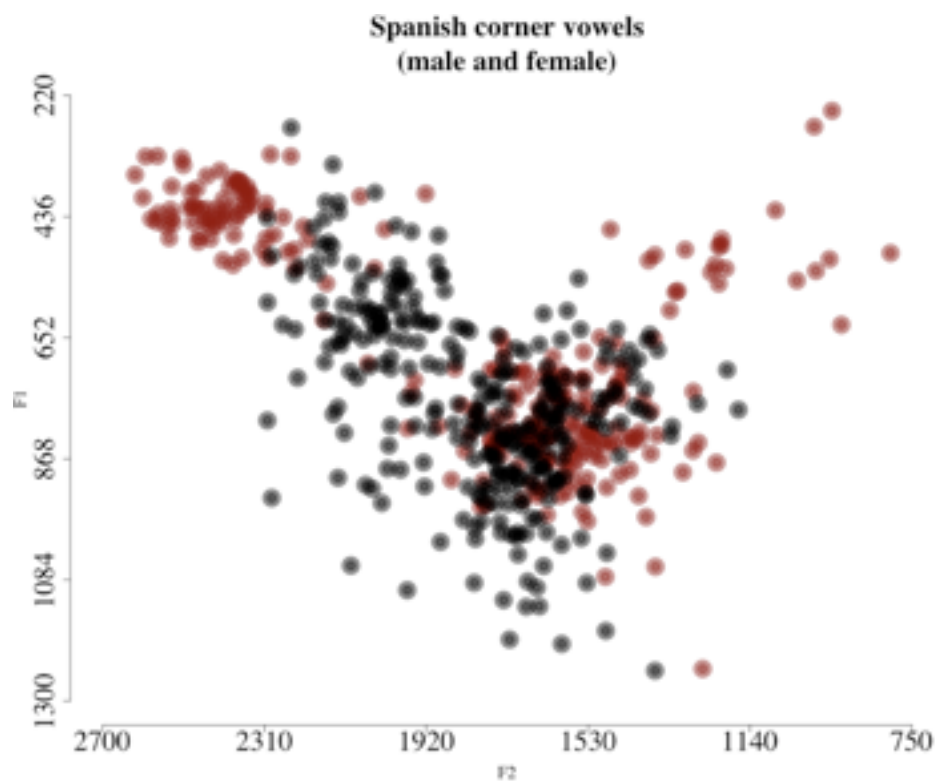
C's and T's

Methods

- Nonparametric Bayesian mixture of Gaussian linear models (Dirichlet process mixture); as many/few categories as the data demands
- 10-fold cross validation
- Fit 10 times, each time holding out 10% of the data for testing on new points
- Fit using MCMC (Gibbs sampler)

C's and T's

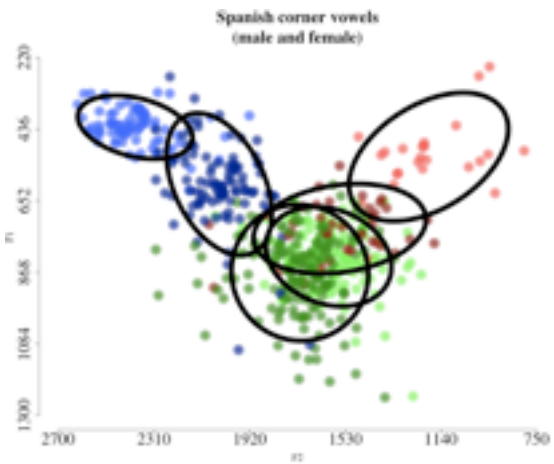
Materials

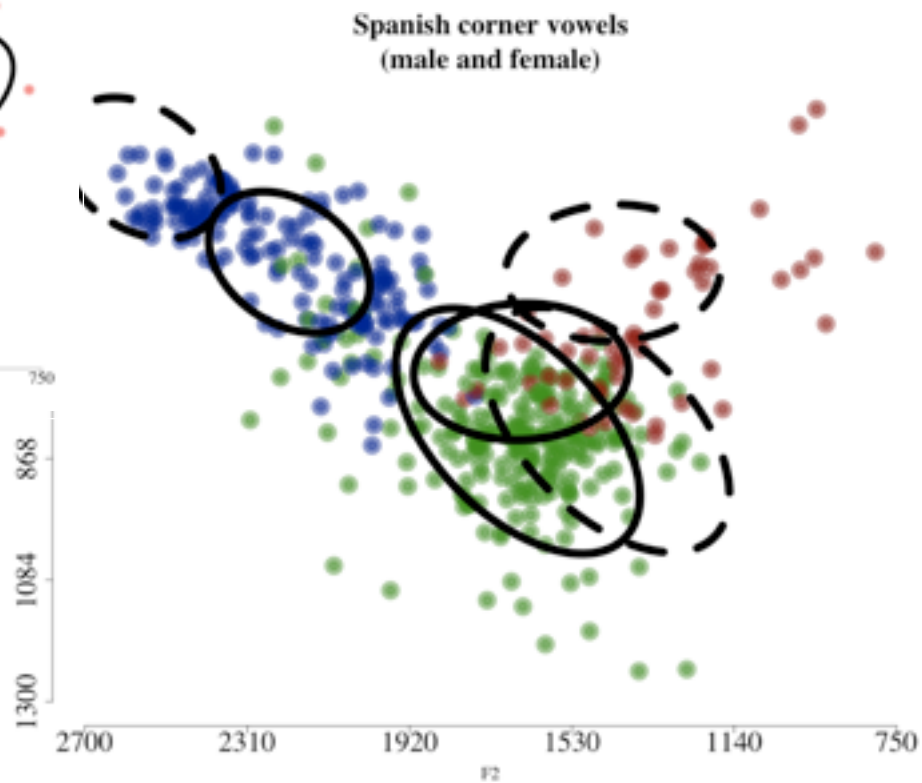


Mark female points as "special"

C's and T's

Results



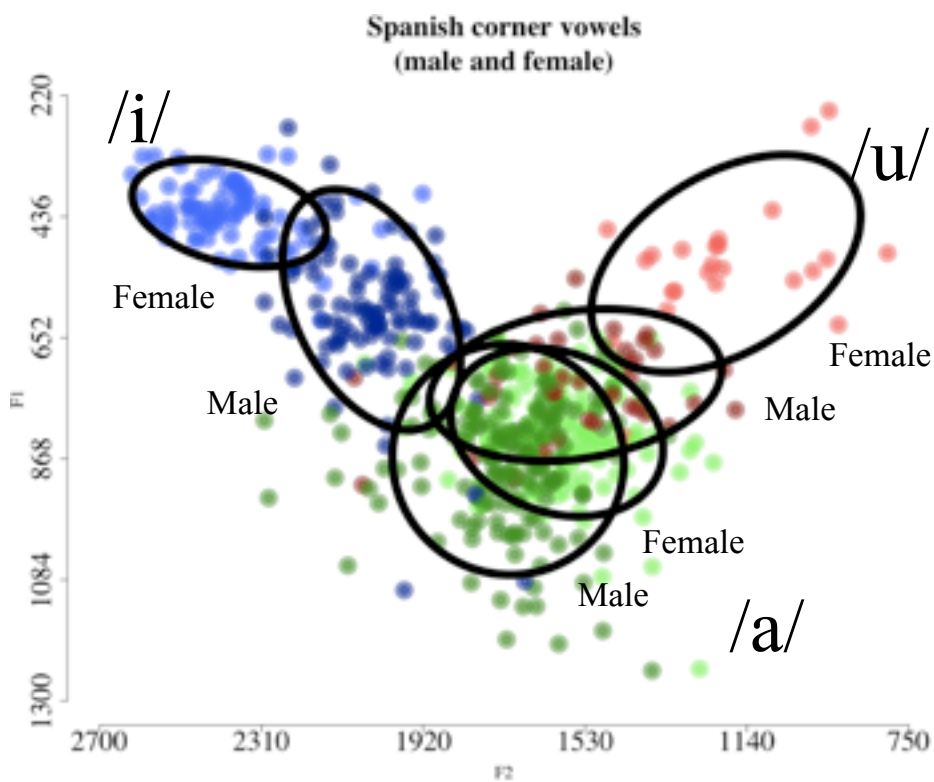


Summary

- It was hard to learn any appropriate categories if you don't know about speaker variability
- New model learns categories by *simultaneously learning categories and sex/speaker-specific transformations*
- Easier to learn appropriate categories if you also learn speaker variability

Labelling?

How do we know which points are "special"?



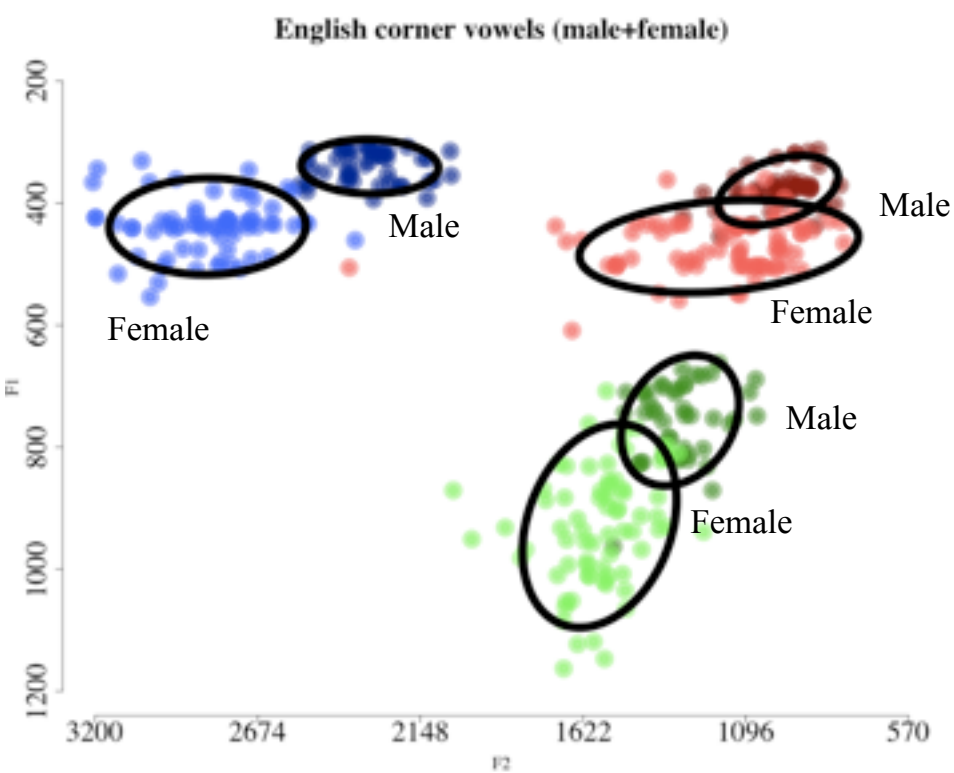
Latent attributes

- **As before:**
  - **Nonparametric Bayesian mixture of Gaussian linear models (Dirichlet process mixture); as many/few categories as the data demands**
- **Now, the same, plus:**
  - **For each point, learn the value of a single bit (either 1 or 0) indicating whether that point is "special"**

Latent attributes

Learning categories + transformations + predictor

Materials



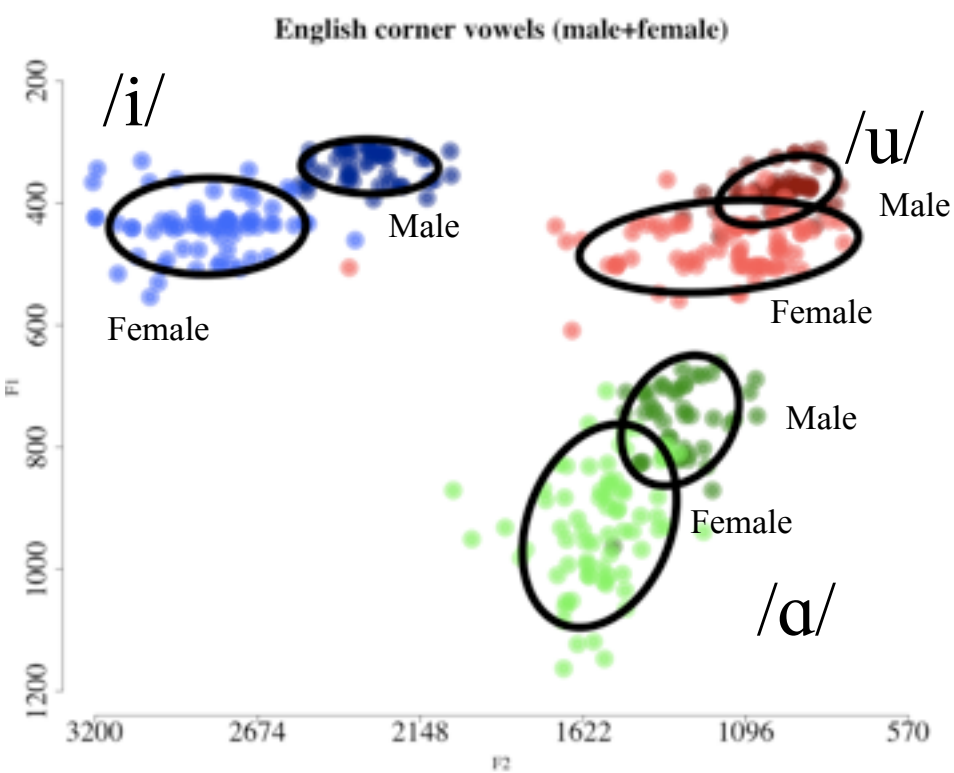
English

- Data from Hillenbrand (1997)
- First four formants measured at steady state from wordlist data
- Corner vowels (/i/, /u/, /a/) extracted as test case
- 344 data points (61% female, 39% male)

Latent attributes

Learning categories + transformations + predictor

Materials



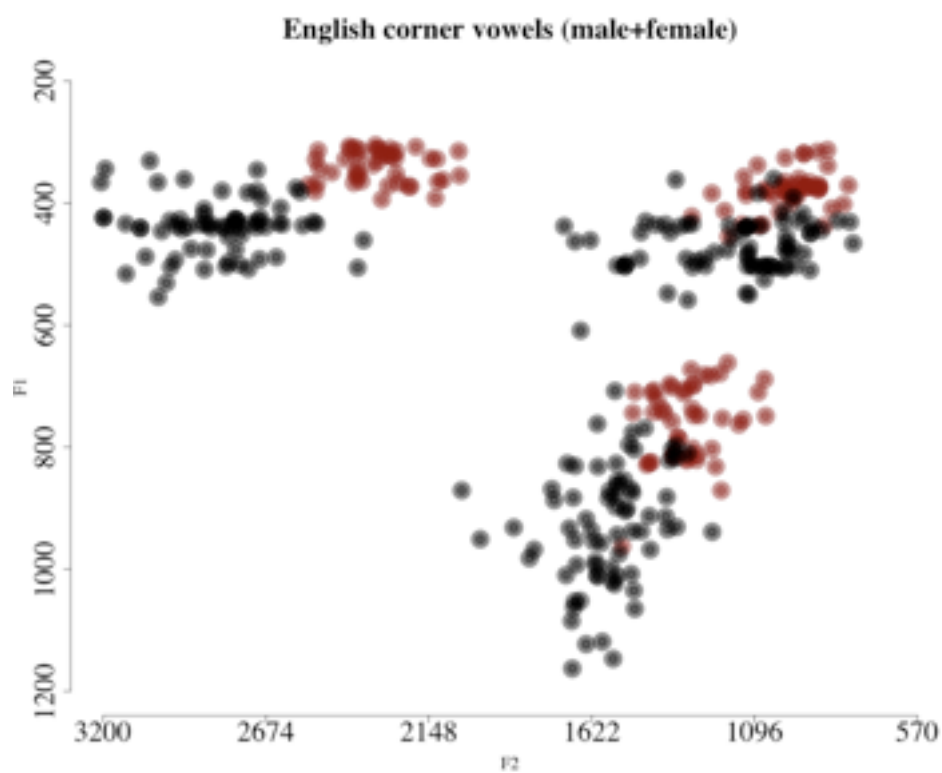
English

- Data from Hillenbrand (1997)
- First four formants measured at steady state from wordlist data
- Corner vowels (/i/, /u/, /a/) extracted as test case
- 344 data points (61% female, 39% male)

Latent attributes

Learning categories + transformations + predictor

Materials



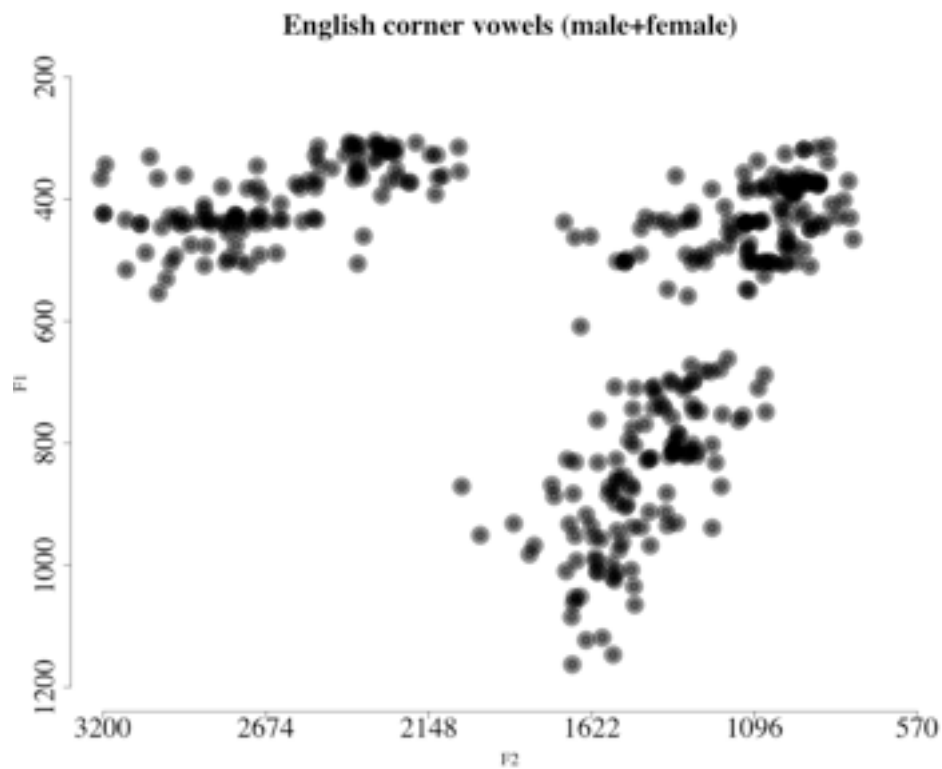
39%
male
61%
female

What we did before

Latent attributes

Learning categories + transformations + predictor

Materials



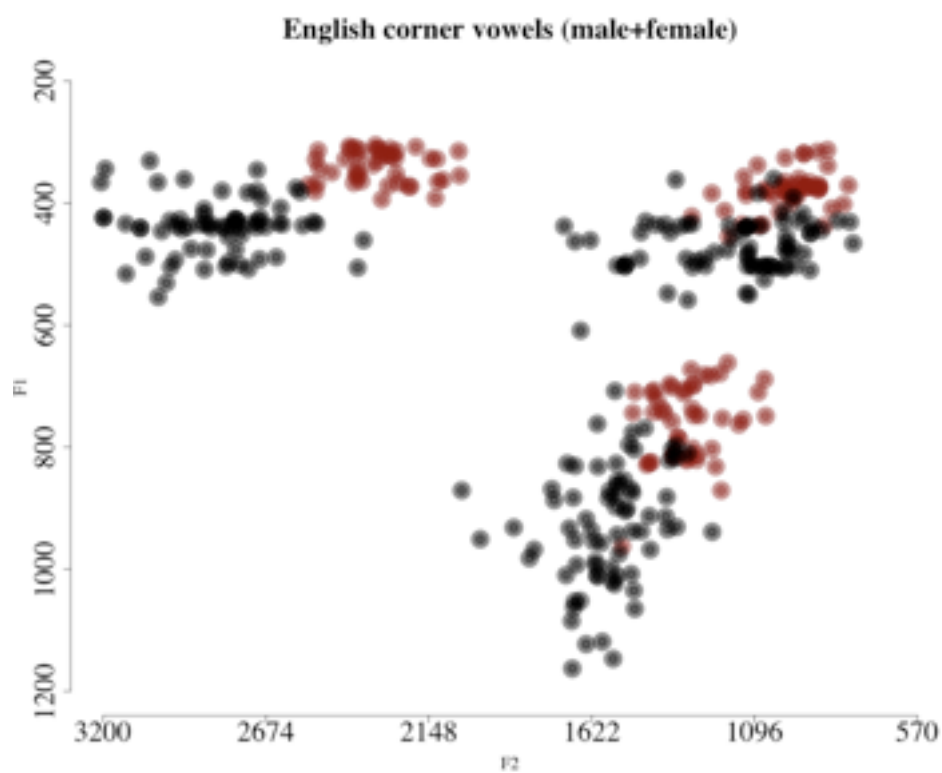
39% male
61% female

What we will do now

Latent attributes

Learning categories + transformations + predictor

Materials



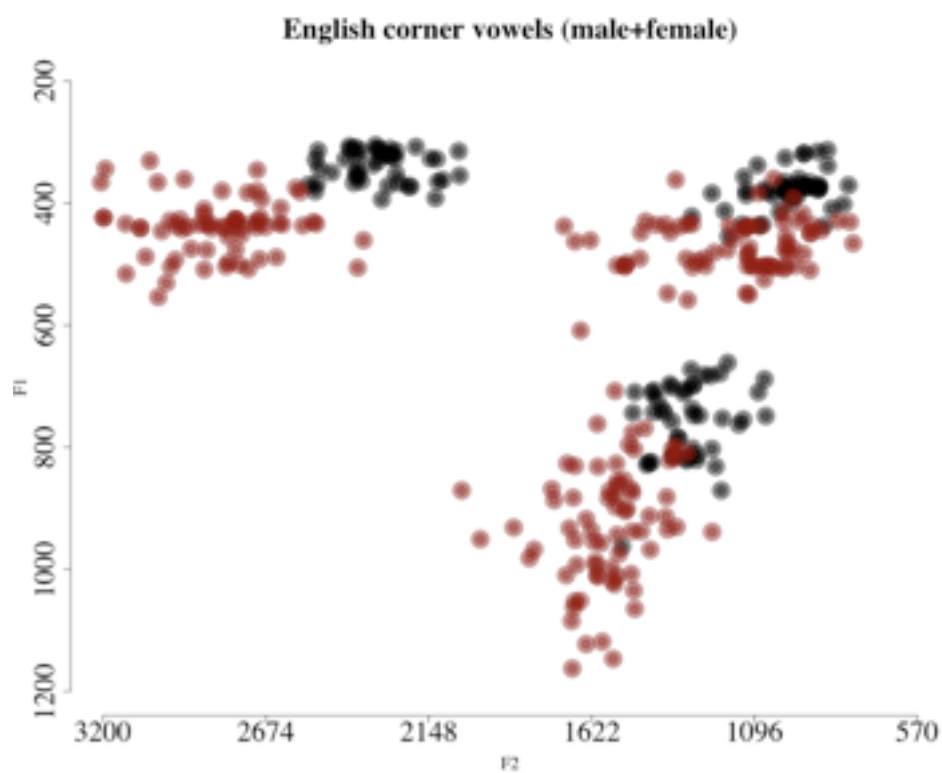
39%
male
61%
female

What we hope to recover

Latent attributes

Learning categories + transformations + predictor

Materials



39%
male
61%
female

What we hope to recover

Latent attributes

Learning categories + transformations + predictor

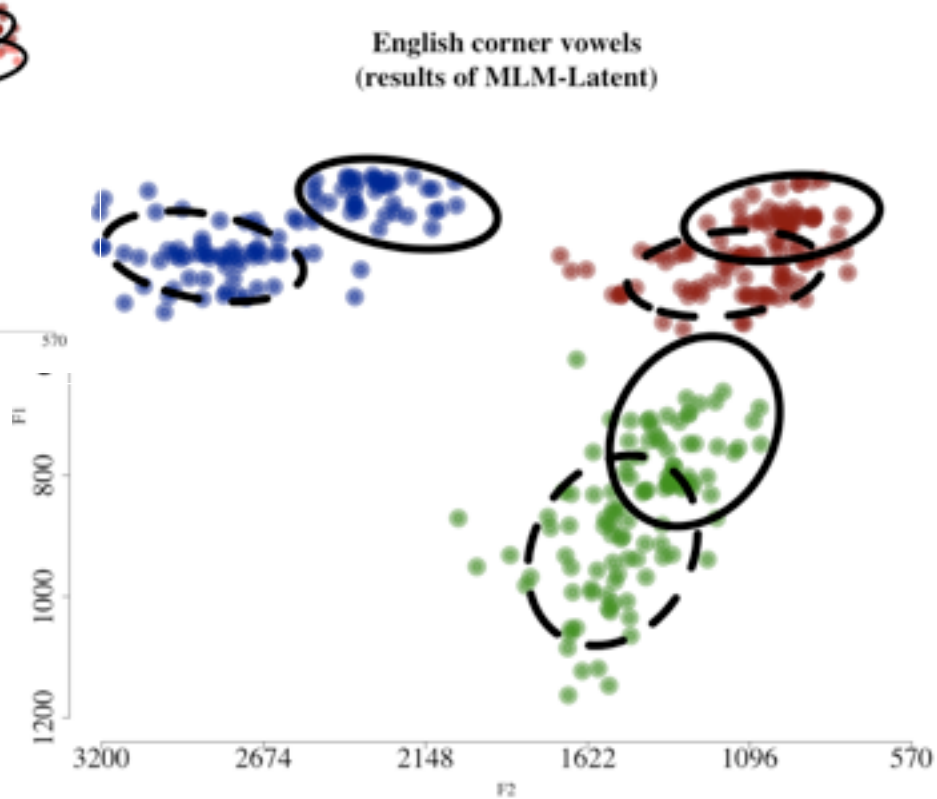
Methods

- Nonparametric Bayesian mixture of Gaussian linear models (Dirichlet process mixture); as many/few categories as the data demands
- 10-fold cross validation
- Fit using MCMC (Gibbs sampler)

Latent attributes

Learning categories + transformations + predictor

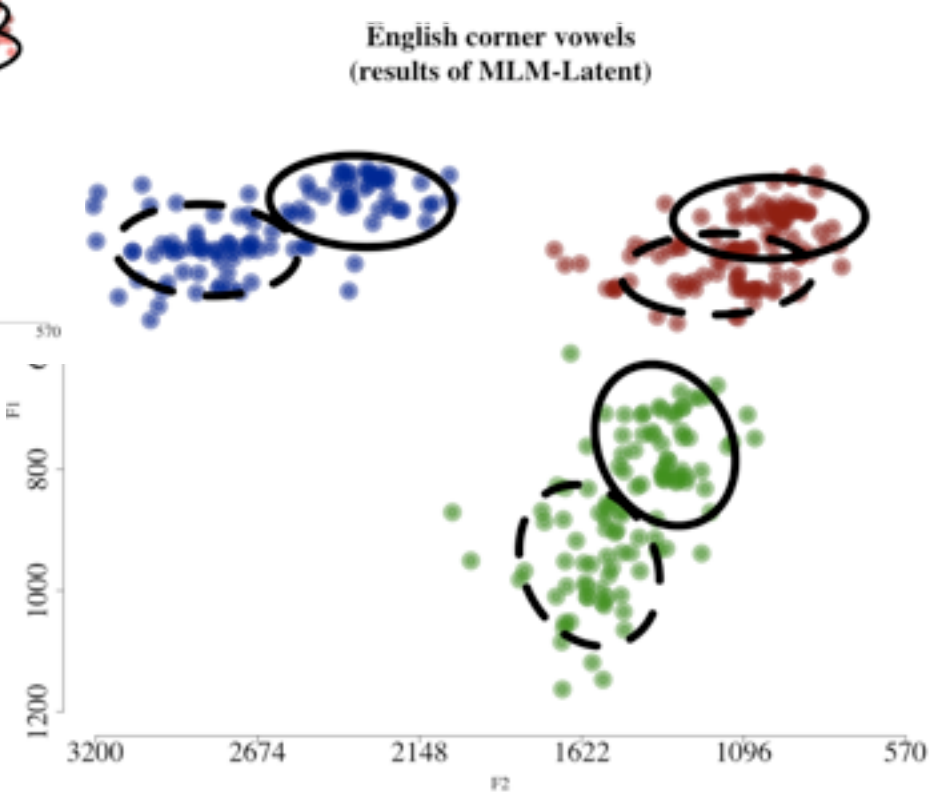
Results



Latent attributes

Learning categories + transformations + predictor

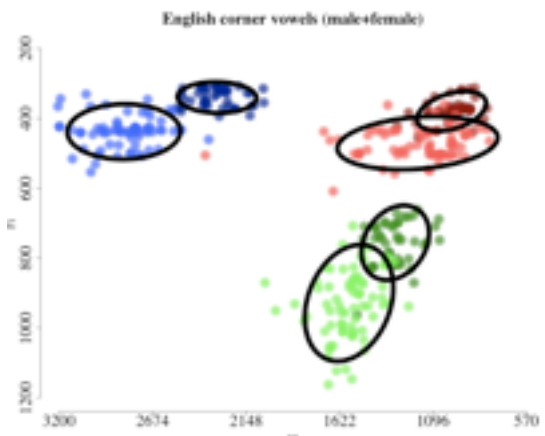
Results

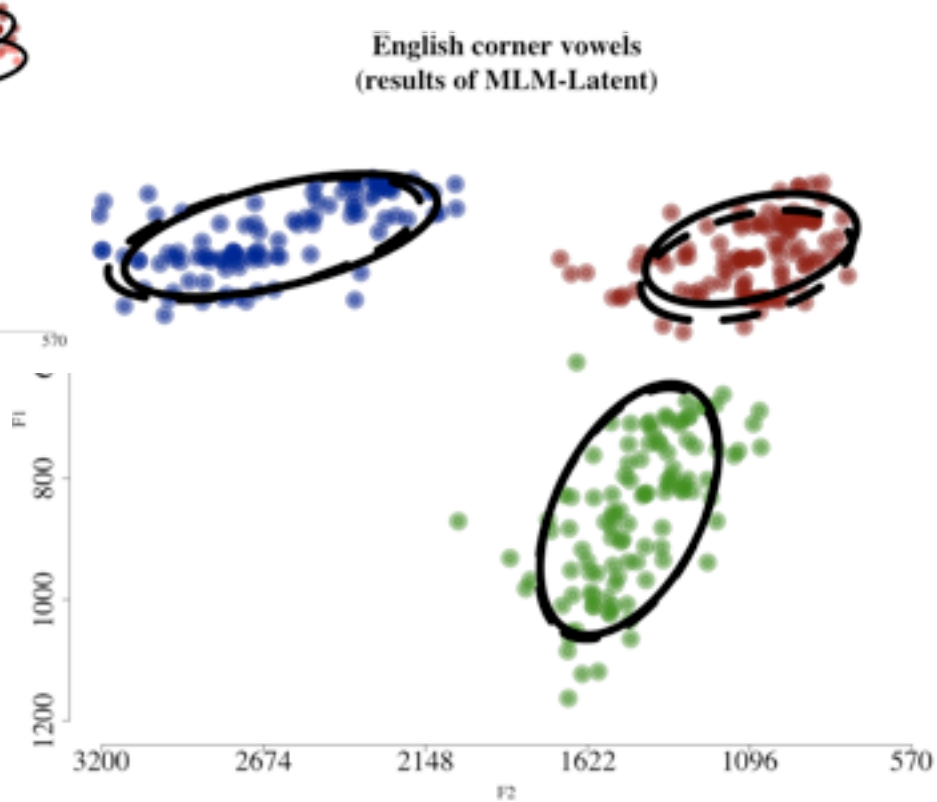


Latent attributes

Learning categories + transformations + predictor

Results

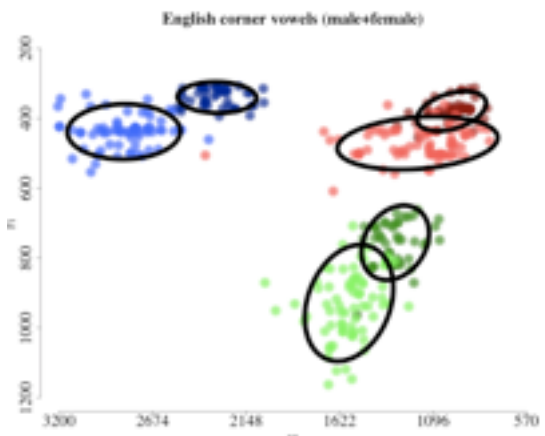


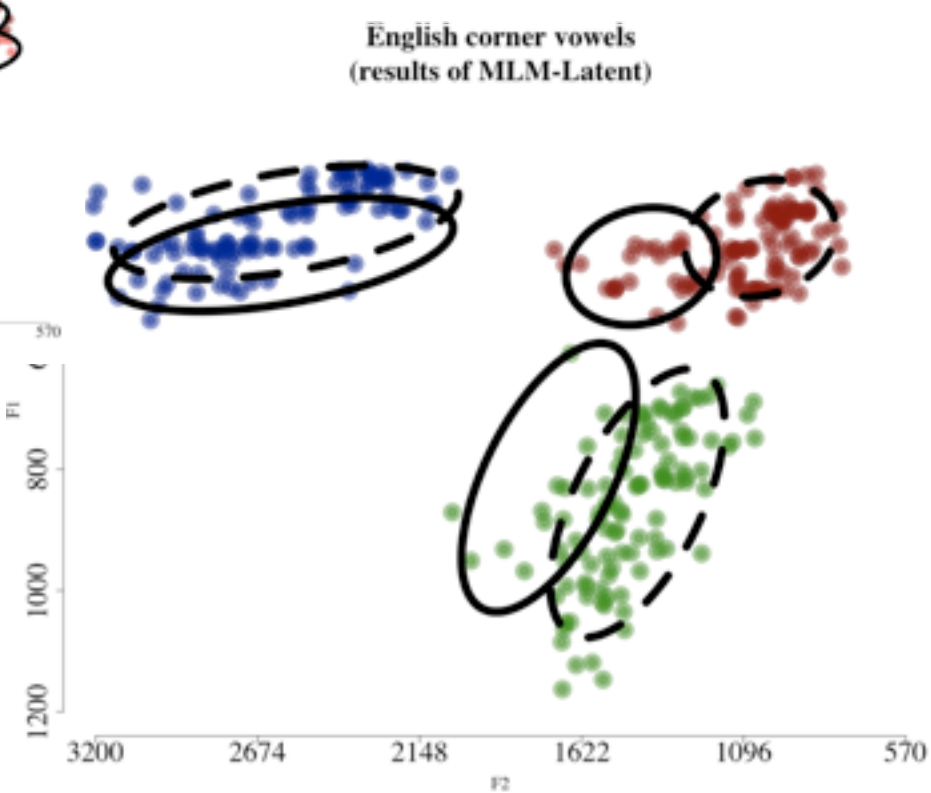


Latent attributes

Learning categories + transformations + predictor

Results





Summary

- Can discover (roughly) which tokens are male/female
- Model searches for categories and shifts and notices phonetically "suspicious" behavior
- Statistical properties of phonetics cue learner to different types of tokens

contextual variants

Inuktitut

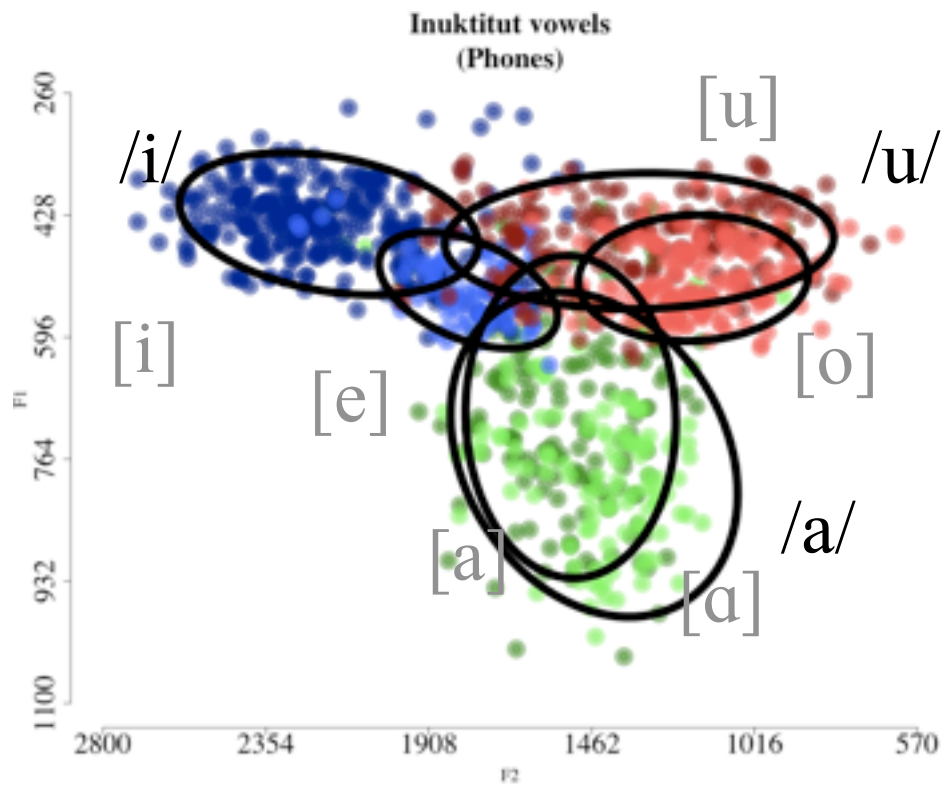
- Eskimo-Aleut, 30,000 speakers
- Three-vowel system
- Uvular consonants cause substantial retraction of all three vowels (i -> e, u -> o, a -> ɑ)
- Easy to find examples of retraction across morpheme boundaries

Inuktitut

- Data from single female speaker from Cape Dorset, Nunavut
- Vowels elicited in word list, formants measured by hand at the center of the vowel

  (Denis and Pollard 2008)
- 239 data points in original corpus, upsampled by jittering to 1000 points

Materials


Inuktitut vowels
(Phones)
/i/
[u]
/u/
[i]
[e]
[o]
/a/
[a]
[ɑ]
260
428
596
764
932
1100
F1
2800
2354
1908
1462
1016
570
F2

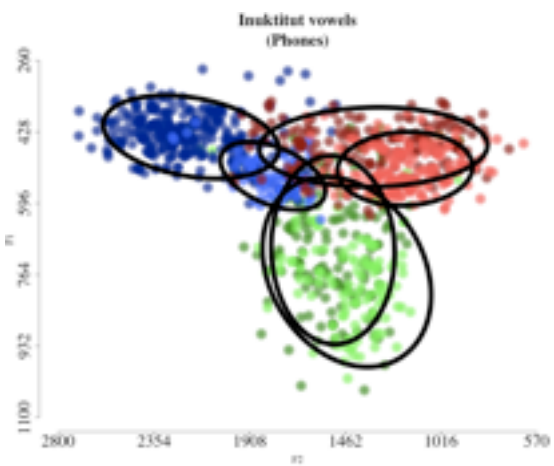
Phones

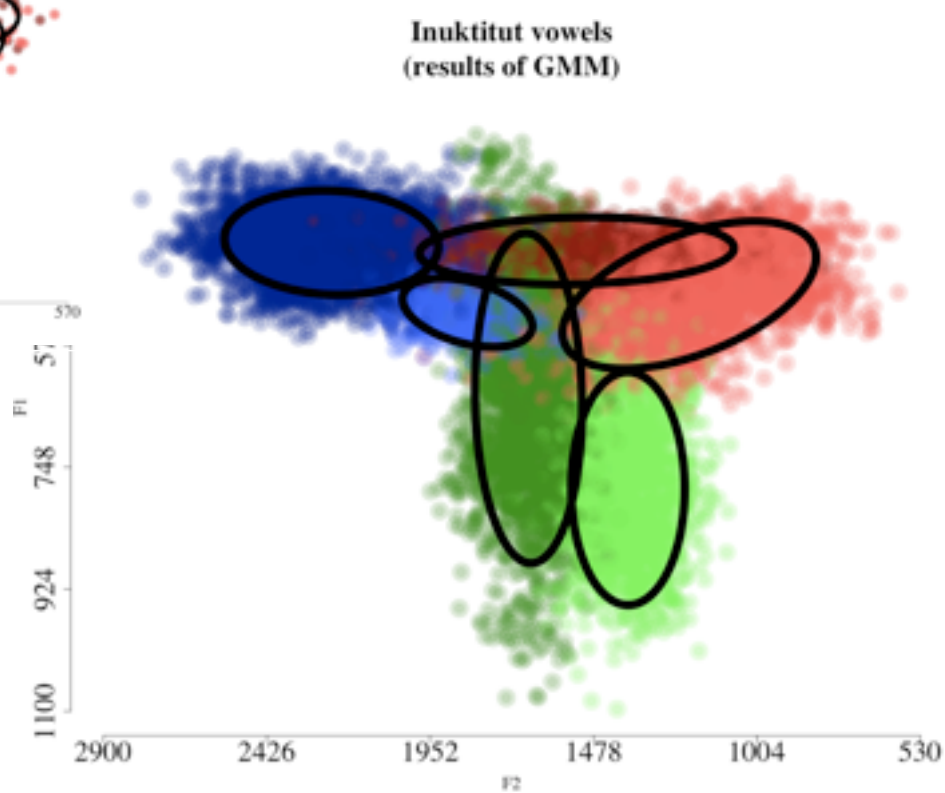
Methods

- Nonparametric Bayesian Gaussian mixture (Dirichlet process mixture); as many/few categories as the data demands
- 10-fold cross validation
- Fit 10 times, each time holding out 10% of the data for testing on new points
- Fit using MCMC (Gibbs sampler)

Phones

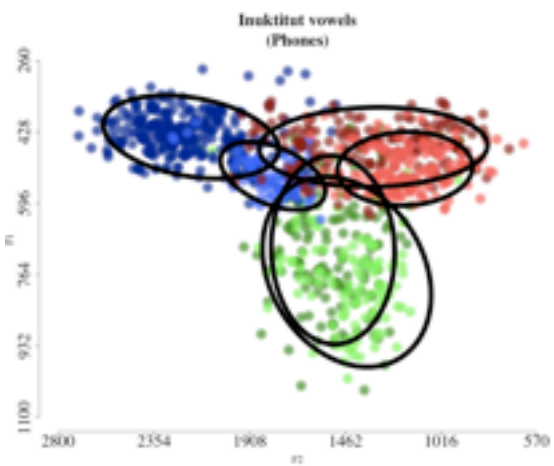
Results

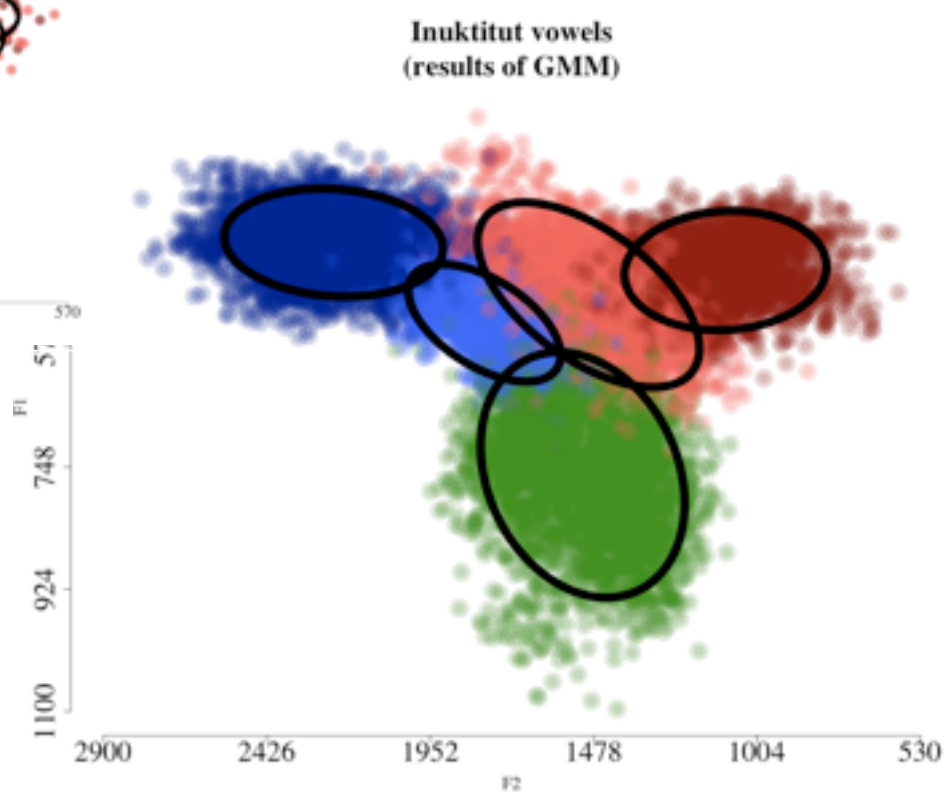




Phones

Results





Group Phones

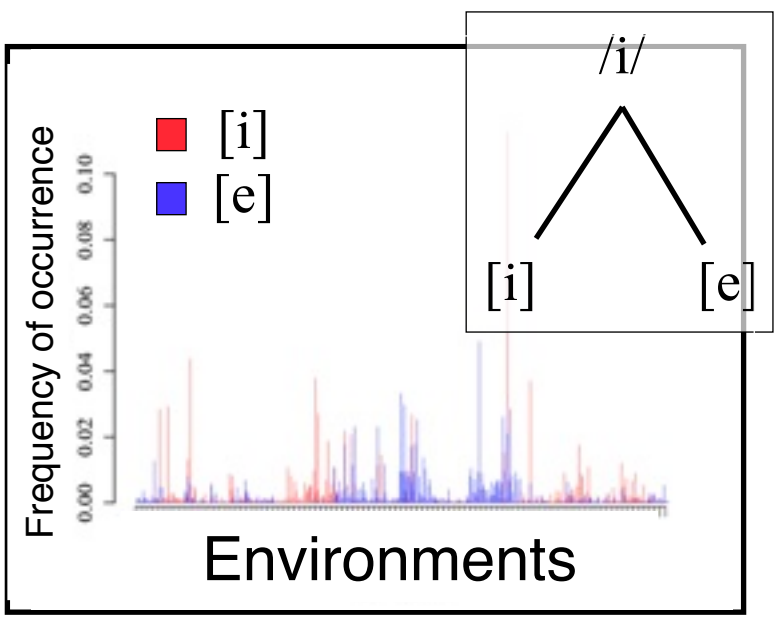
Learning phonemes

Methods

- Examine models and manually map Gaussian categories to nearest Inuktitut phones
- Use Peperkamp et al.'s (2006) statistical method for grouping phones into phonemes

Group phones

Methods



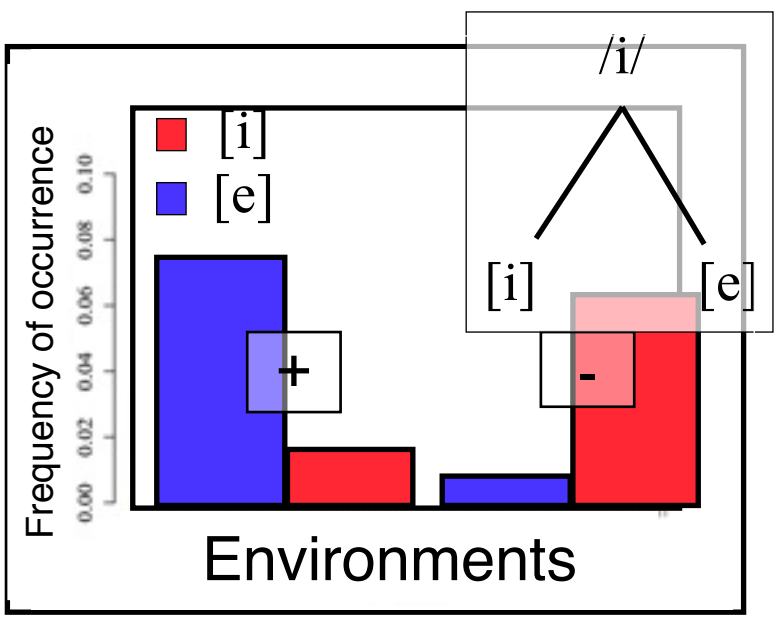
(Peperkamp et al. 2006)

KL divergence

- Number representing how different two probability distributions are
- In this case, probability is "following uvular or not"
- Compare for each pair of phonetic categories found in Experiment 3a

Group phones

Methods



(Peperkamp et al. 2006)

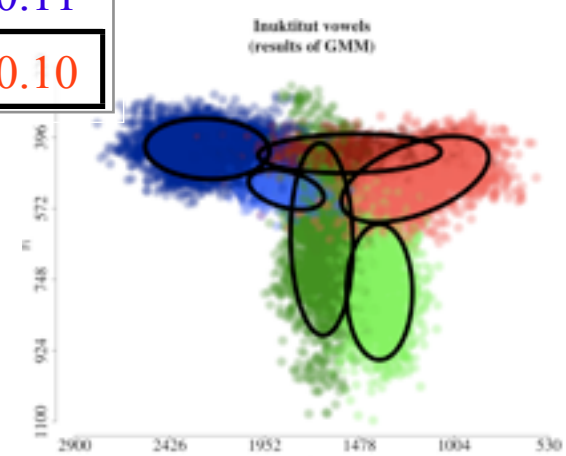
KL divergence

- Number representing how different two probability distributions are
- In this case, probability is "following uvular or not"
- Compare for each pair of phonetic categories found in Experiment 3a

KL divergence

Results

	[i]	[e]	[u]	[o]	[a]	[ɑ]
[i]	0	0.81	0.03	0.32	0.33	0.85
[e]	0.81	0	0.48	0.10	0.09	0.00
[u]	0.03	0.48	0	0.14	0.14	0.50
[o]	0.32	0.10	0.14	0	0.00	0.11
[a]	0.33	0.09	0.14	0.00	0	0.10



Summary

- Learning lexical inventories by relating allophones requires that we first learn the allophones as a surface inventory
- Learning from real data is messy
- Problems learning surface inventories undermine correct learning of lexical inventories (downstream contamination)

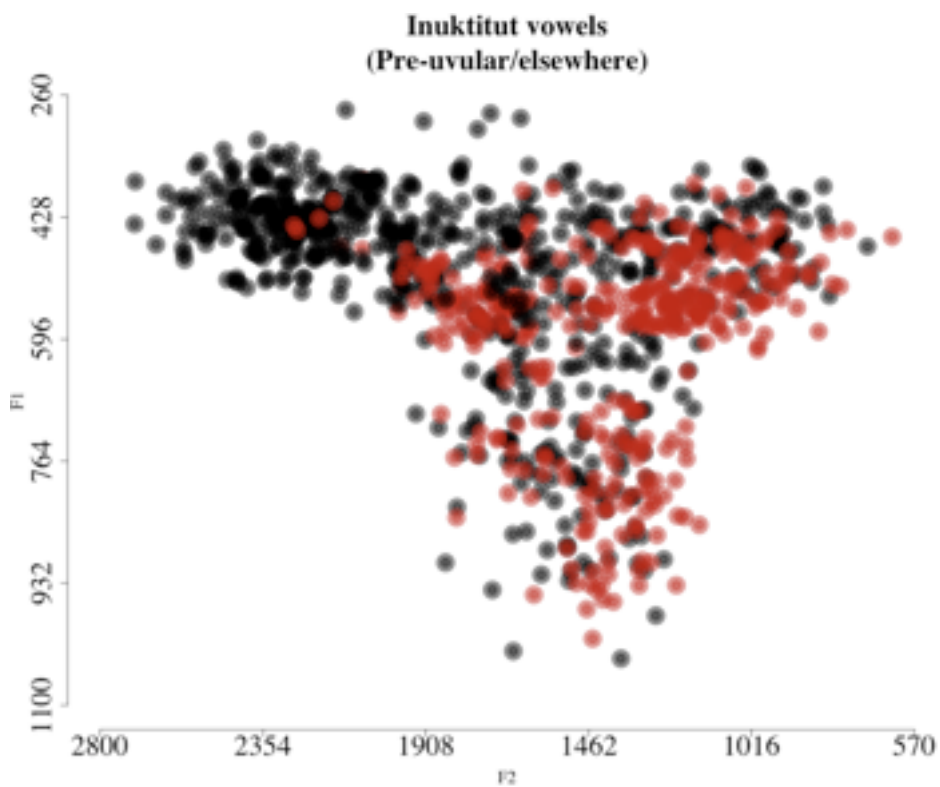
C's and T's

Methods

- Nonparametric Bayesian mixture of Gaussian linear models (Dirichlet process mixture); as many/few categories as the data demands
- 10-fold cross validation
- Fit 10 times, each time holding out 10% of the data for testing on new points
- Fit using MCMC (Gibbs sampler)

C's and T's

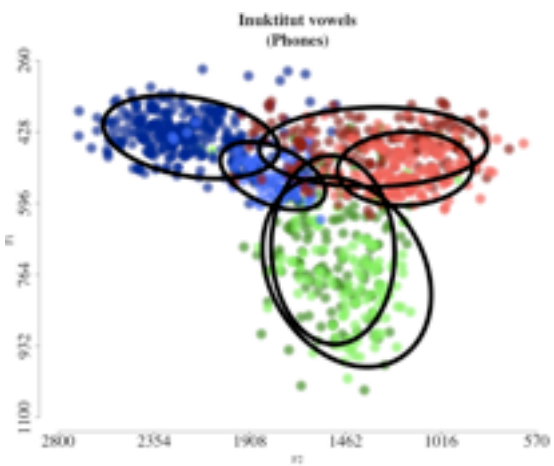
Materials

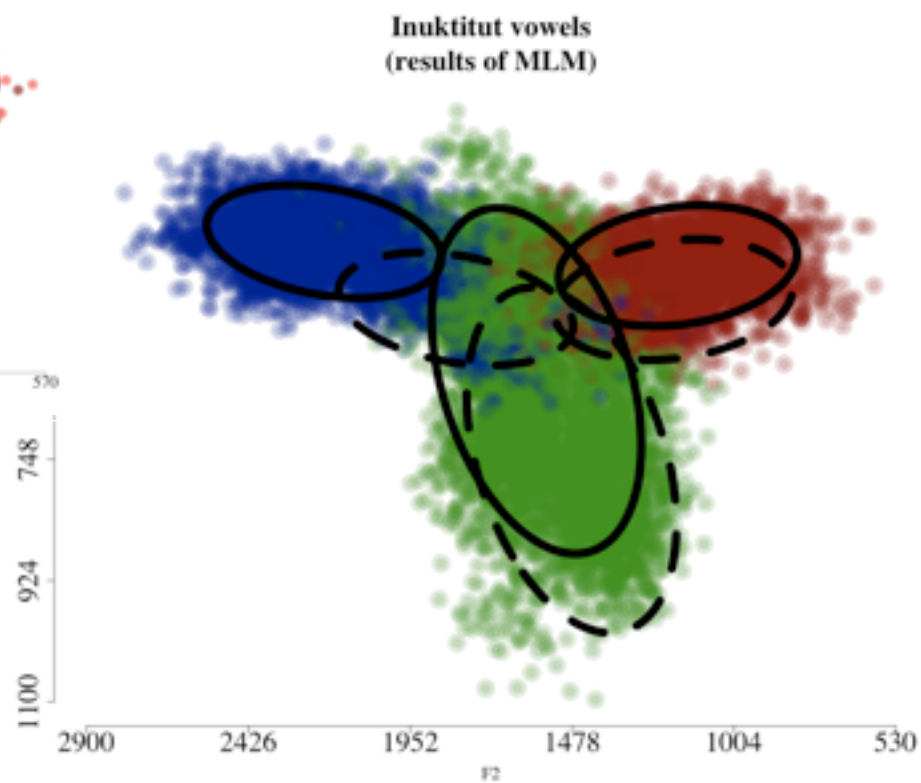


Mark pre-uvular points as "special"

C's and T's

Results



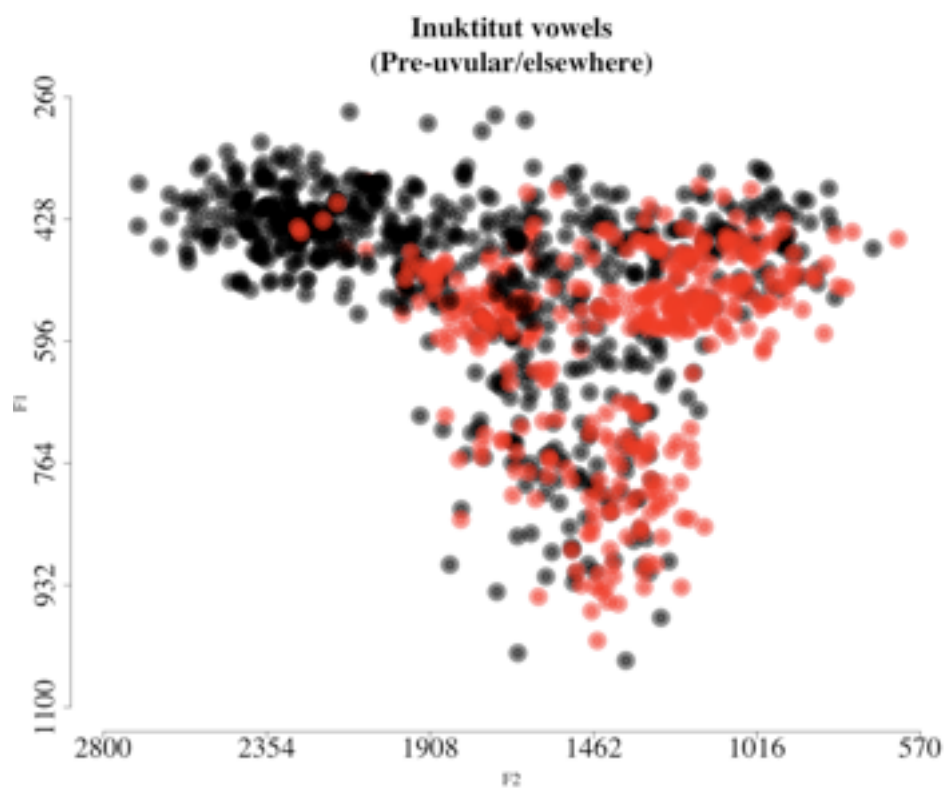


Summary

- New model learns abstract phonetic categories by *simultaneously learning phonetic categories and transformations*
- Succeeds at learning correct abstract categories where a learner which works by finding and grouping phones fails
- Same learning model handles speaker variability and allophonic variability

C’s and T’s and contexts

Materials

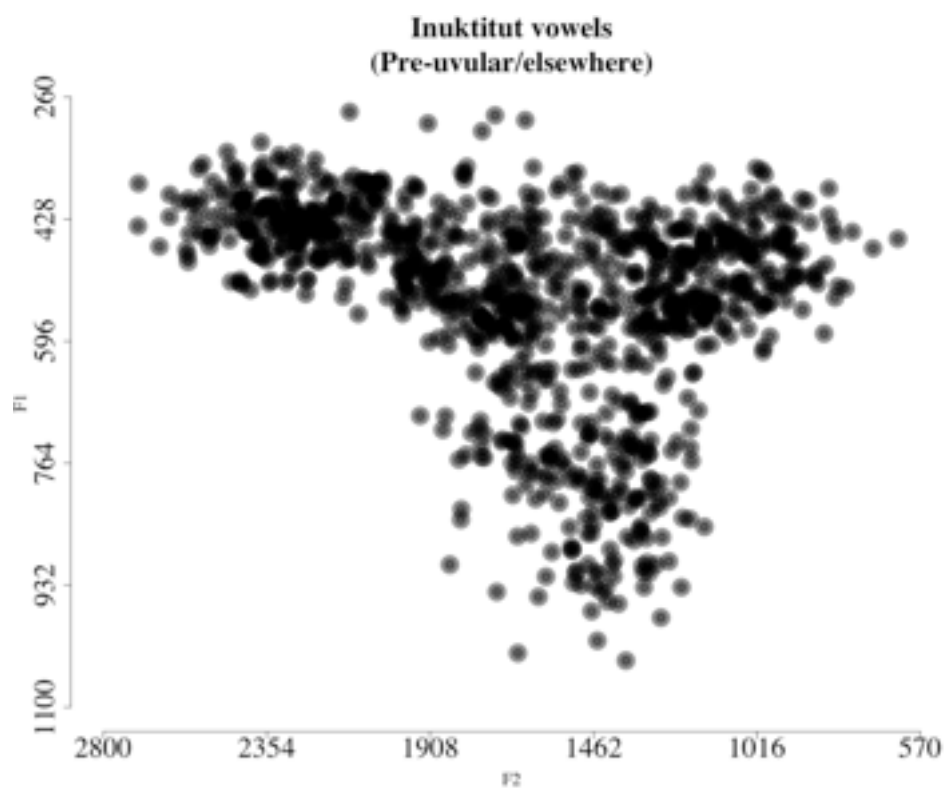


36%
pre-uvular
64%
“elsewhere”

What we did before

C’s and T’s and contexts

Materials

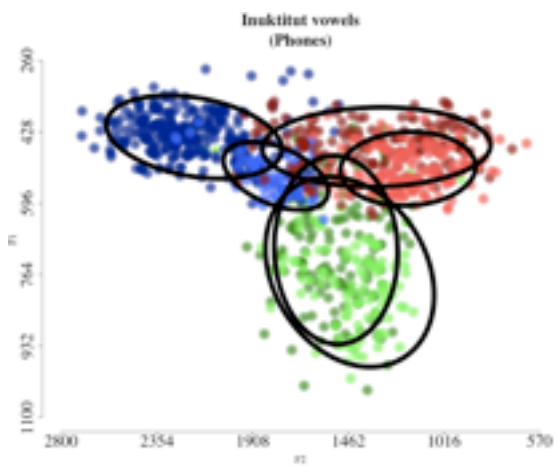


36%
pre-uvular
64%
“elsewhere”

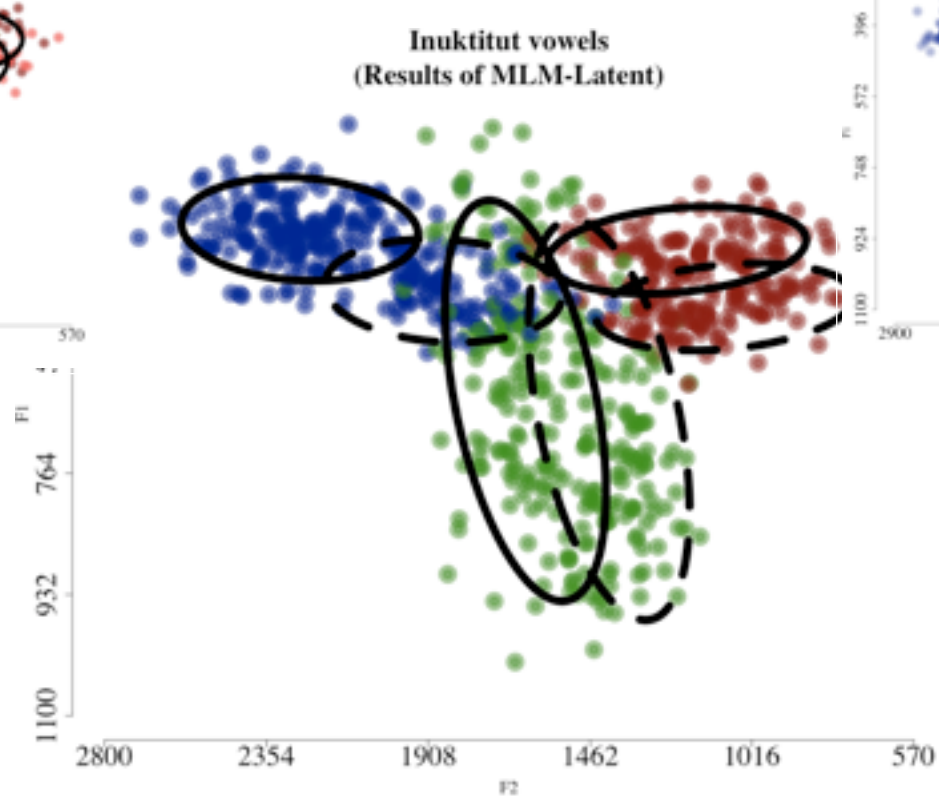
What we will do now

C’s and T’s and contexts

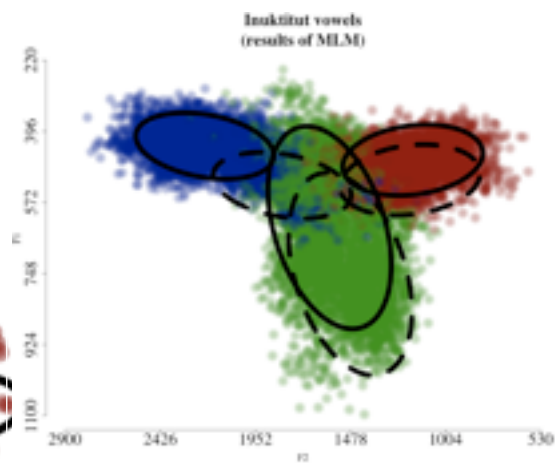
Ground truth



Results

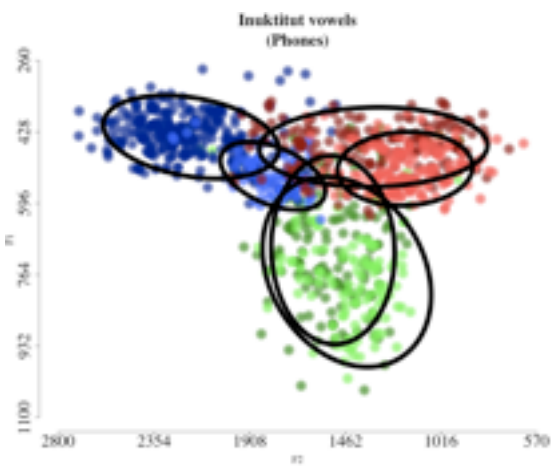


Previous result

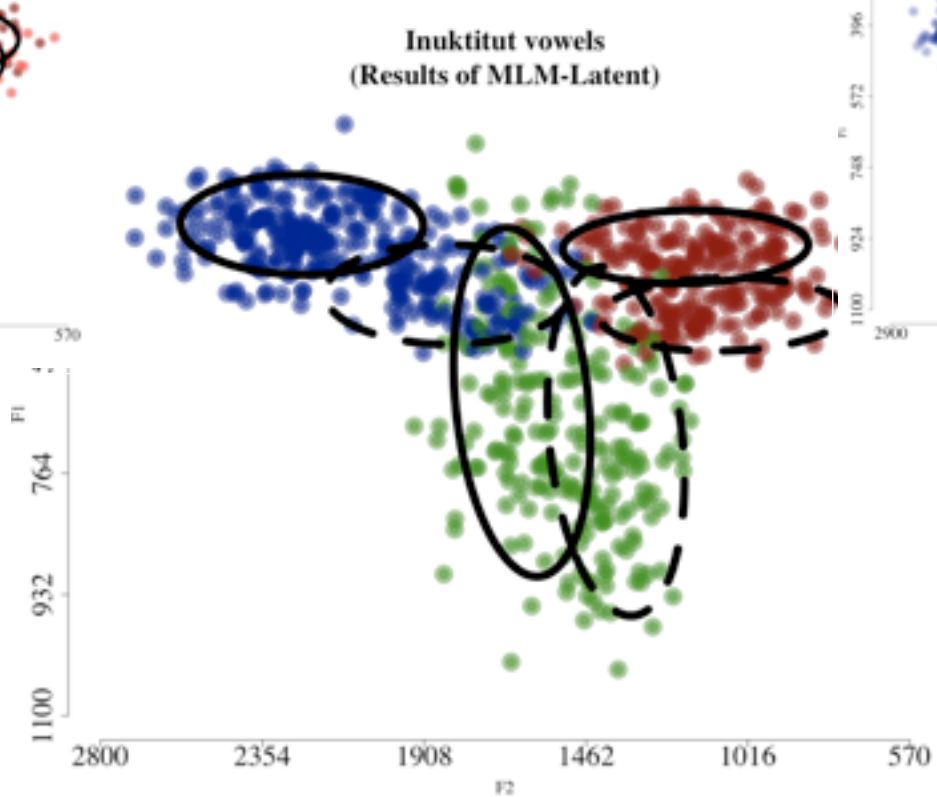


C’s and T’s and contexts

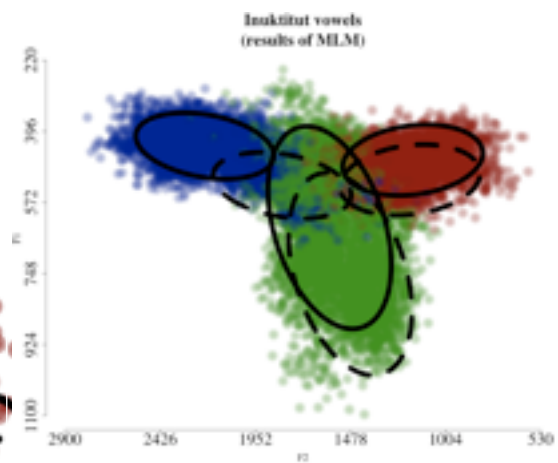
Ground truth



Results

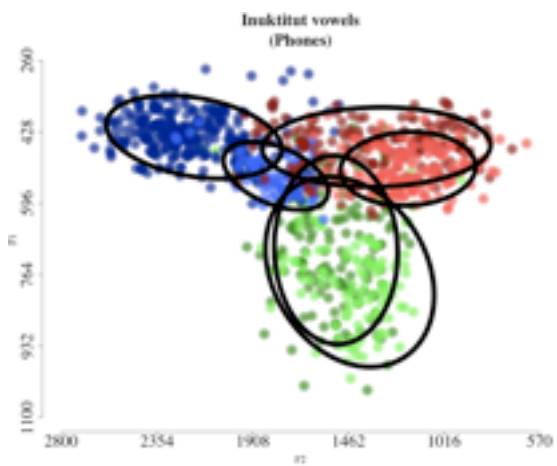


Previous result

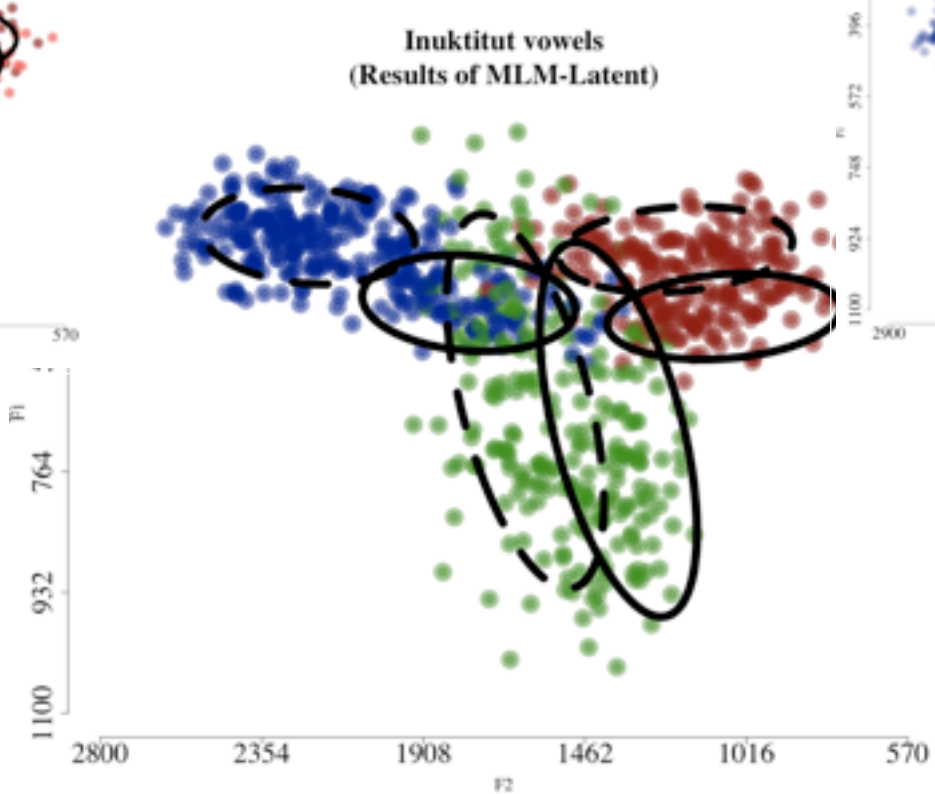


C’s and T’s and contexts

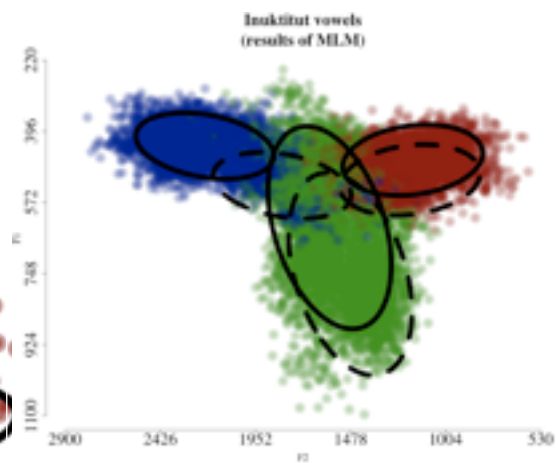
Ground truth



Results

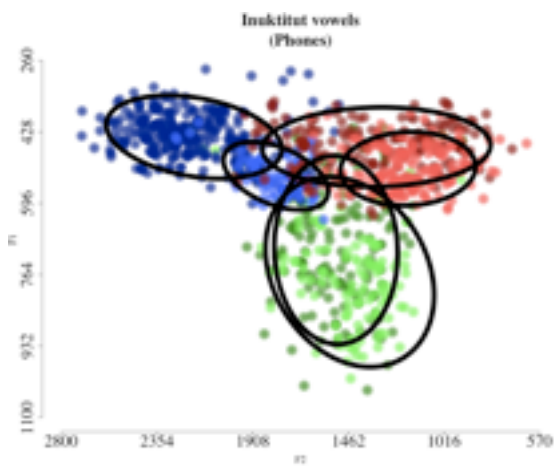


Previous result

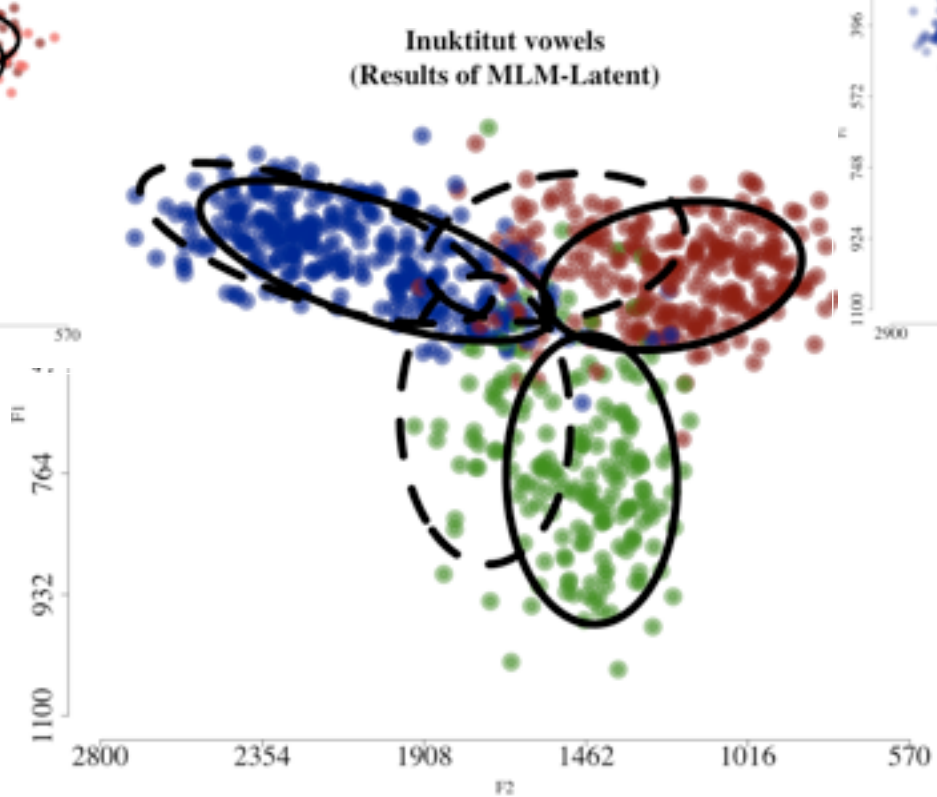


C’s and T’s and contexts

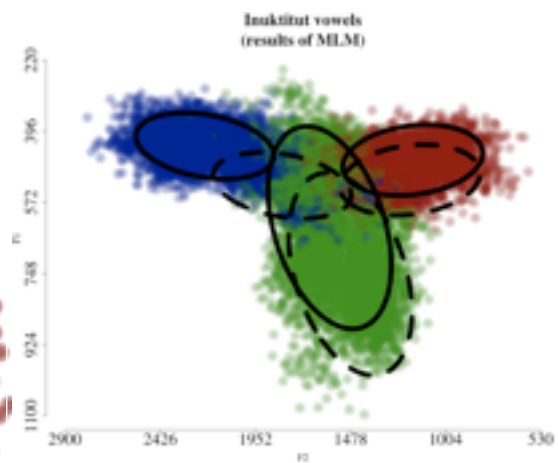
Ground truth



Results



Previous result



Summary

- Can find an allophonic phonetic rule without knowing anything about the environment (almost)
- Model searches for categories and shifts and notices phonetically "suspicious" behavior
- Phonetics cues learner to different types of tokens

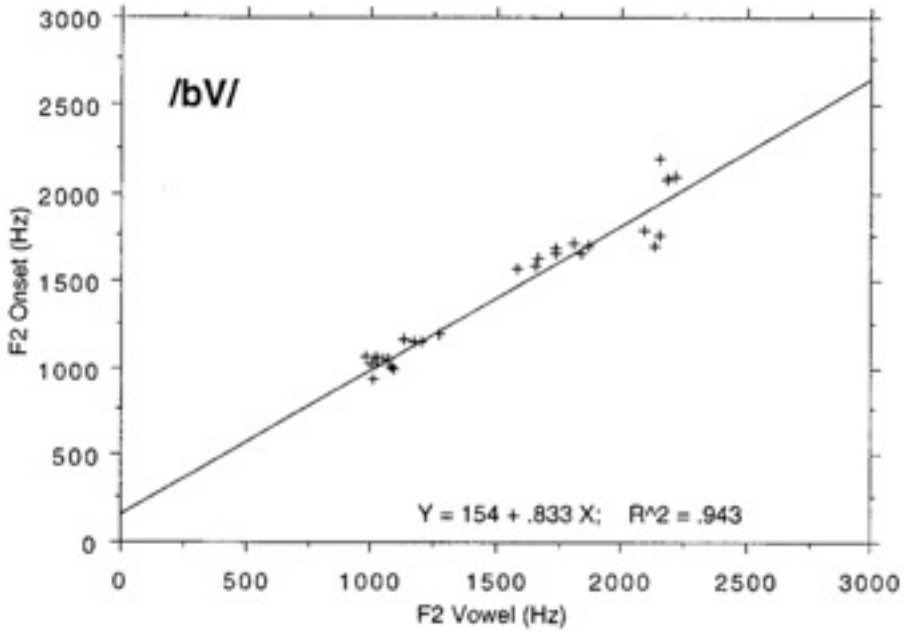
higher order invariants

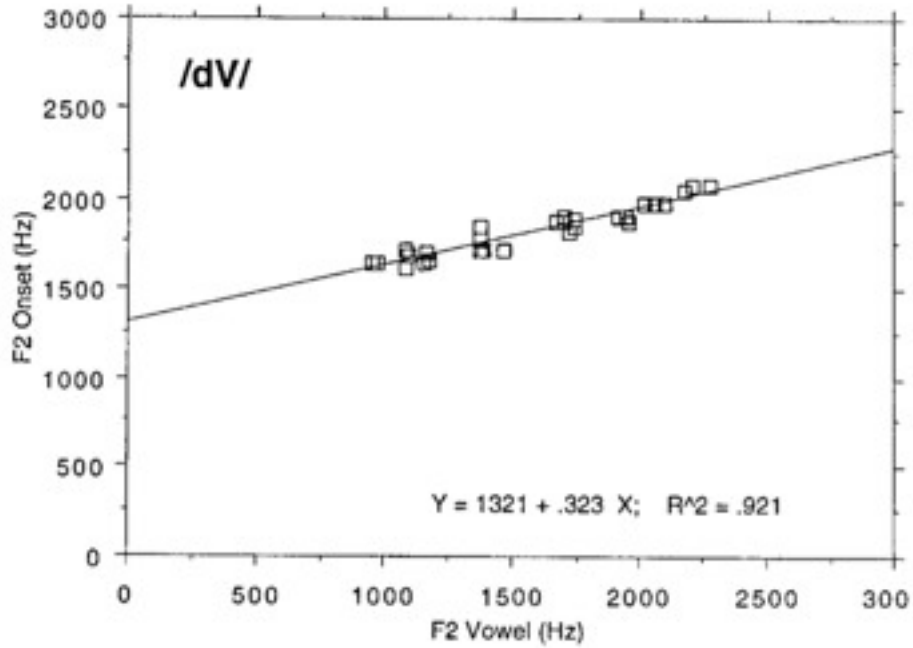
Stevens, etc.

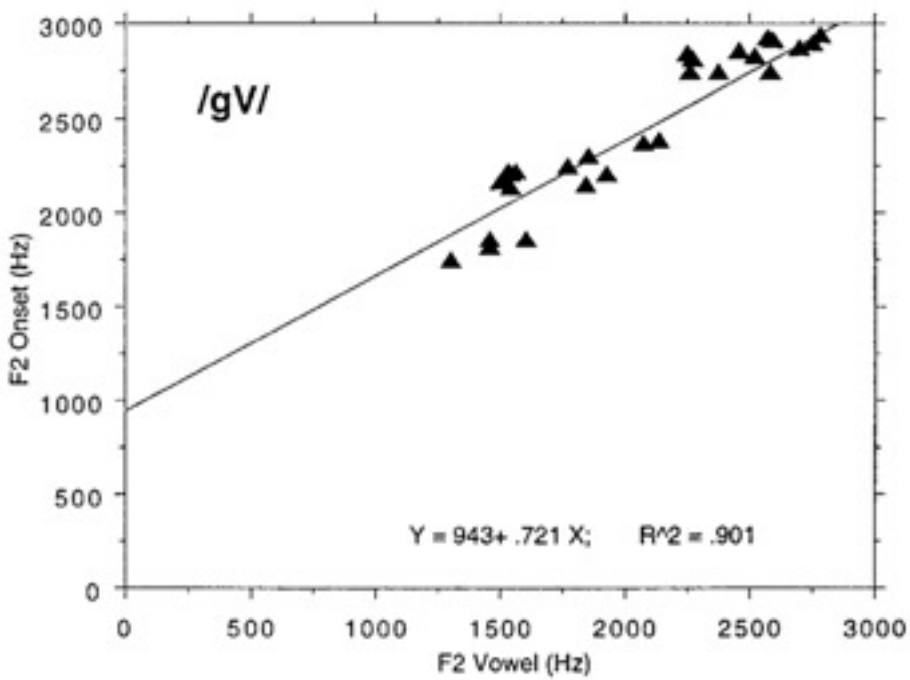
- one possible solution to the problem of variability is to look for derived (higher-order) quantities that are more stable
- features

Locus equations

- Lindblom, Sussman
- consonant place classes correlate with the change in F2 between vowel onset and vowel midpoint







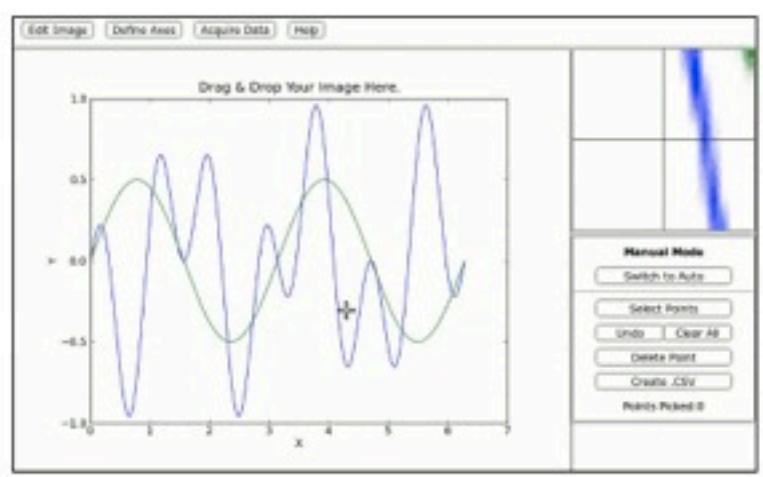
Locus Equations (Sussman)

WebPlotDigitizer

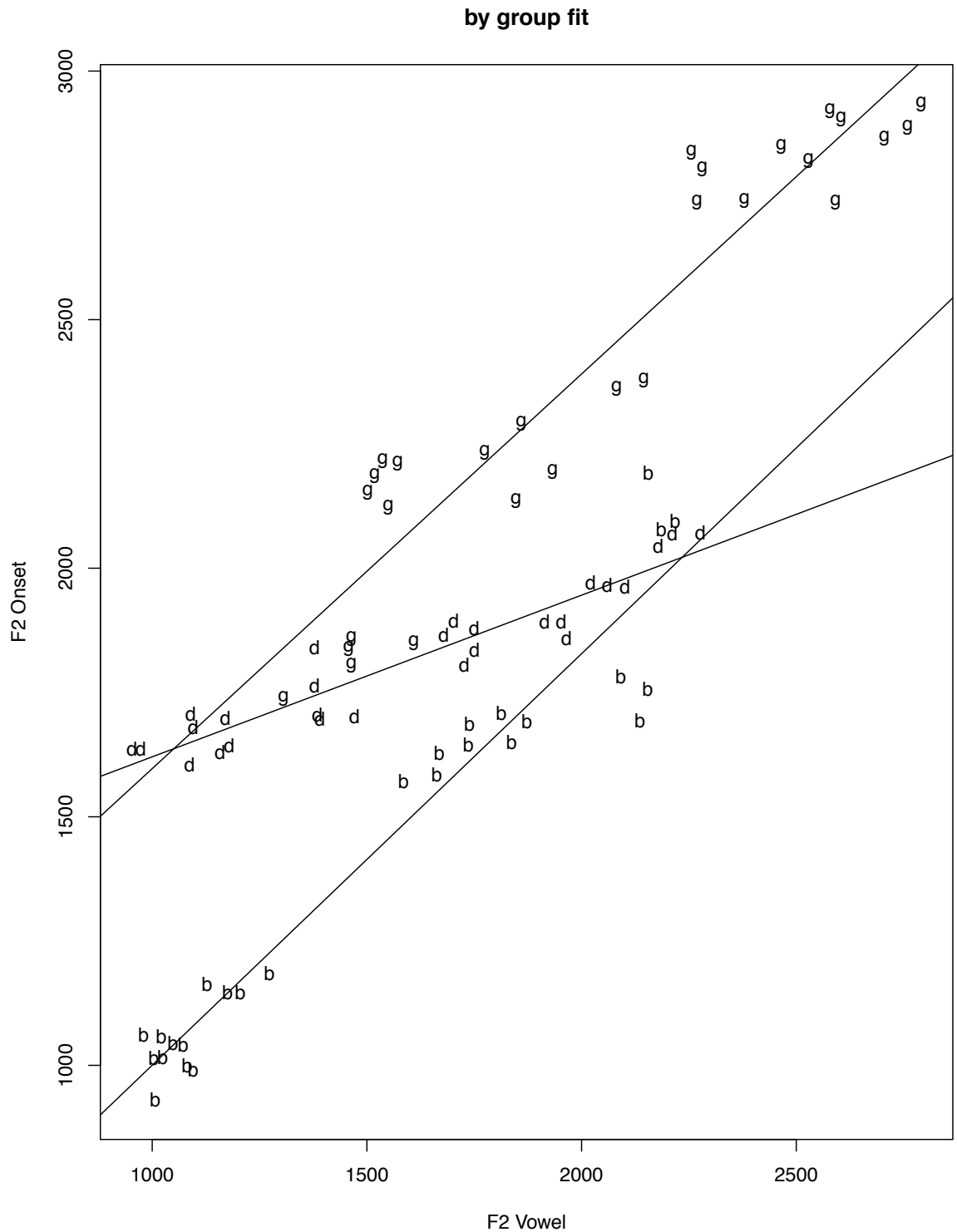
A large quantity of published data is available only in the form of plots and it is often difficult to extract numerical data accurately out of these pictures. There are several softwares available to aid this process, but most are either paid or poorly written. Also, most of the existing programs require Microsoft Windows to work and support only 2D X-Y plots. Due to these limitations WebPlotDigitizer has been developed to facilitate easy and accurate data extraction from a variety of plot types and also maps. This program is built using HTML5 which allows it to run within a browser and requires no installation on to the user's hard drive.

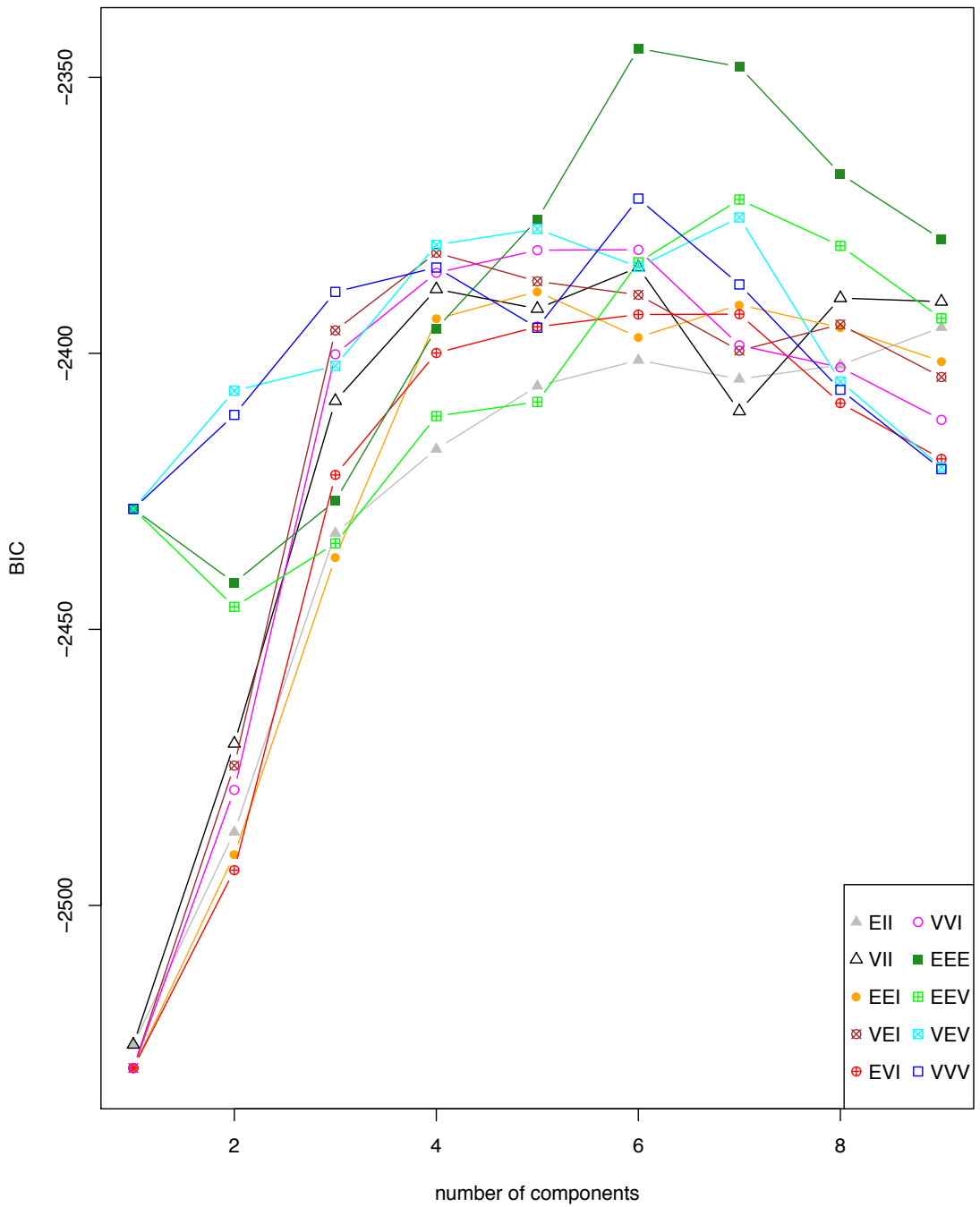
Version 2.4 Released

- Web based. No installation needed. Just point your browser to the launch button and start working!
- Drag and Drop any plot image directly to the screen.
- A zoomed up view on the side aids accurate selection of data points.
- Generates data in .CSV format which can be used by any data analysis program like Excel, OpenOffice, Origin etc.
- Supports XY charts (even skewed and non-orthogonal), polar plots, ternary diagrams and maps.
- Automatic curve extraction algorithms aid rapid extraction of large number of points.
- This program is free of charge. The project is distributed under the GNU General Public License Version 3.

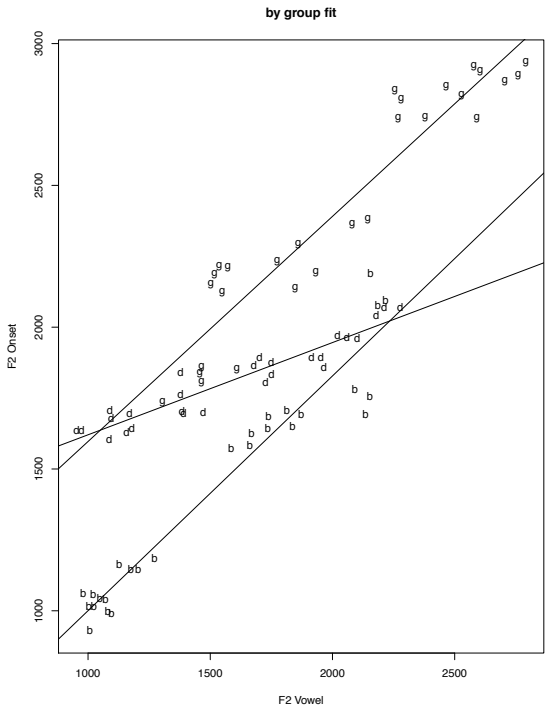


Ground "truth": 3 regressions



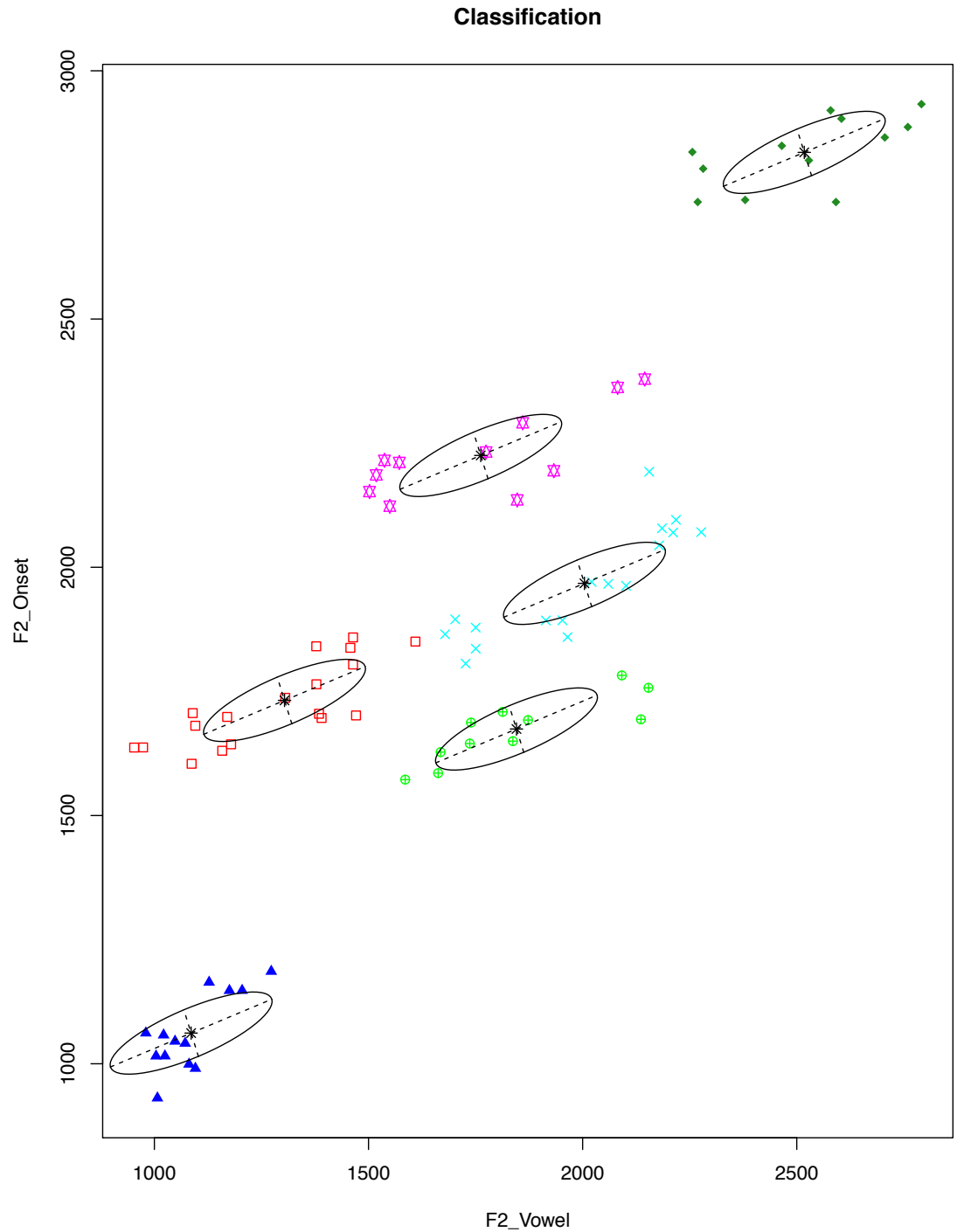


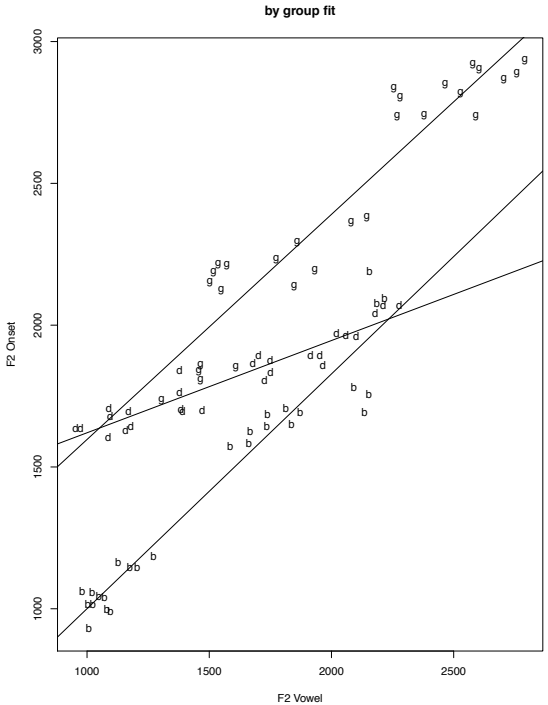
EEE = elliposidal, equal volume, shape and orientation



Gaussian Mixture Model (mclust)

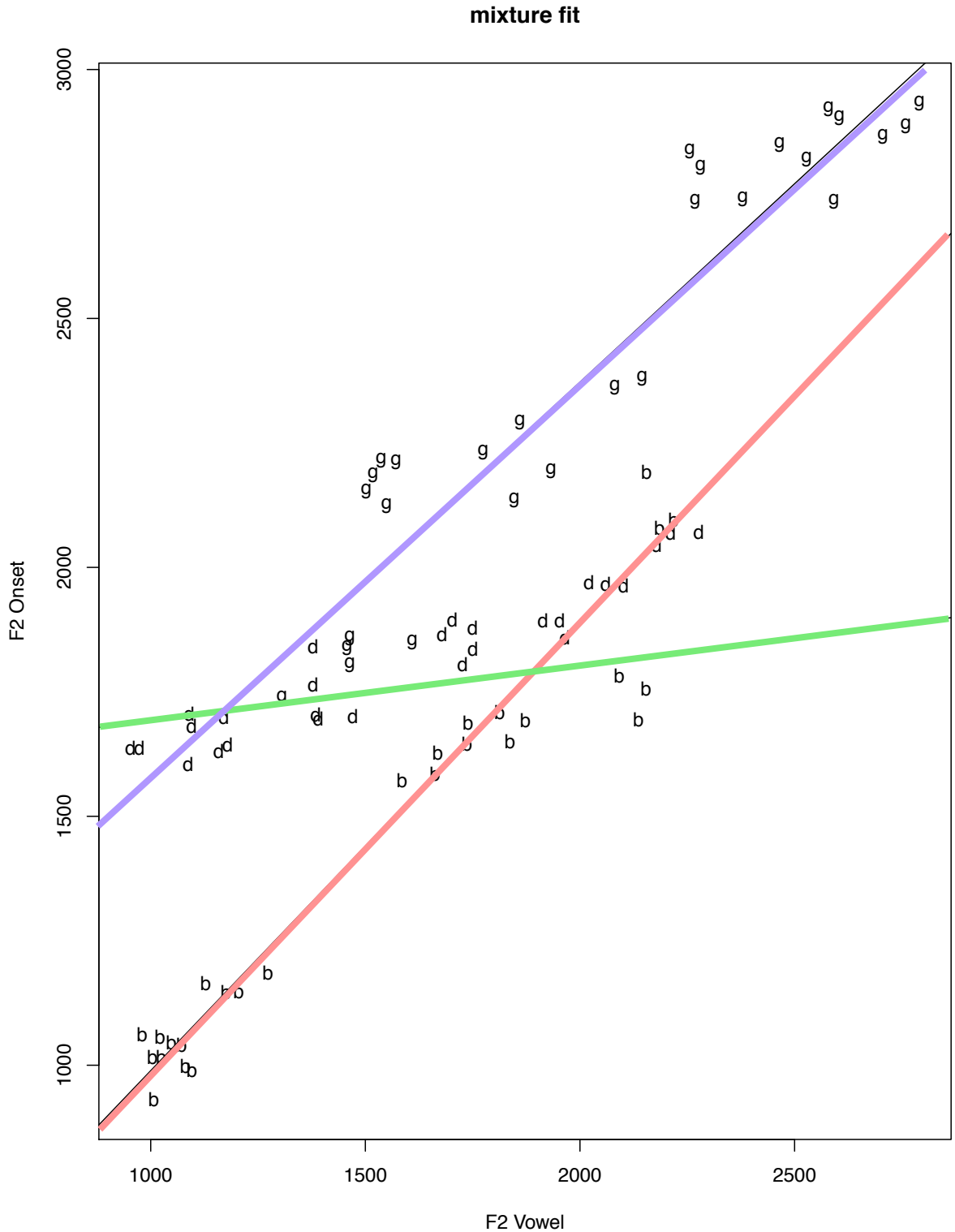
best model (BIC): *6 categories*

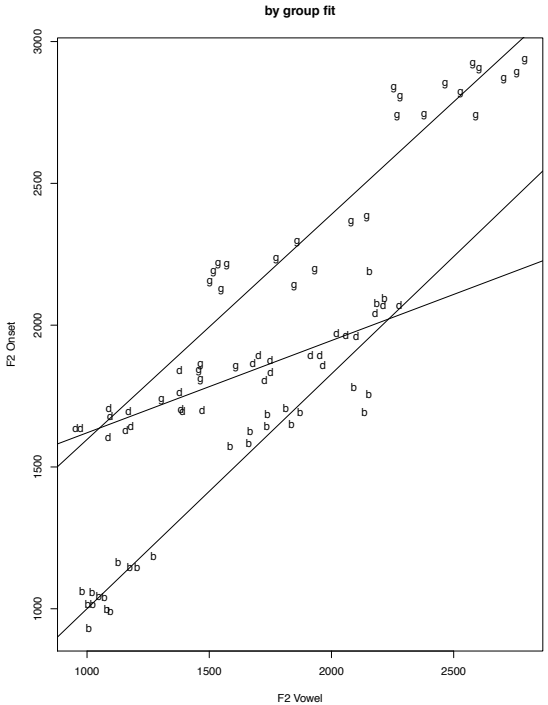




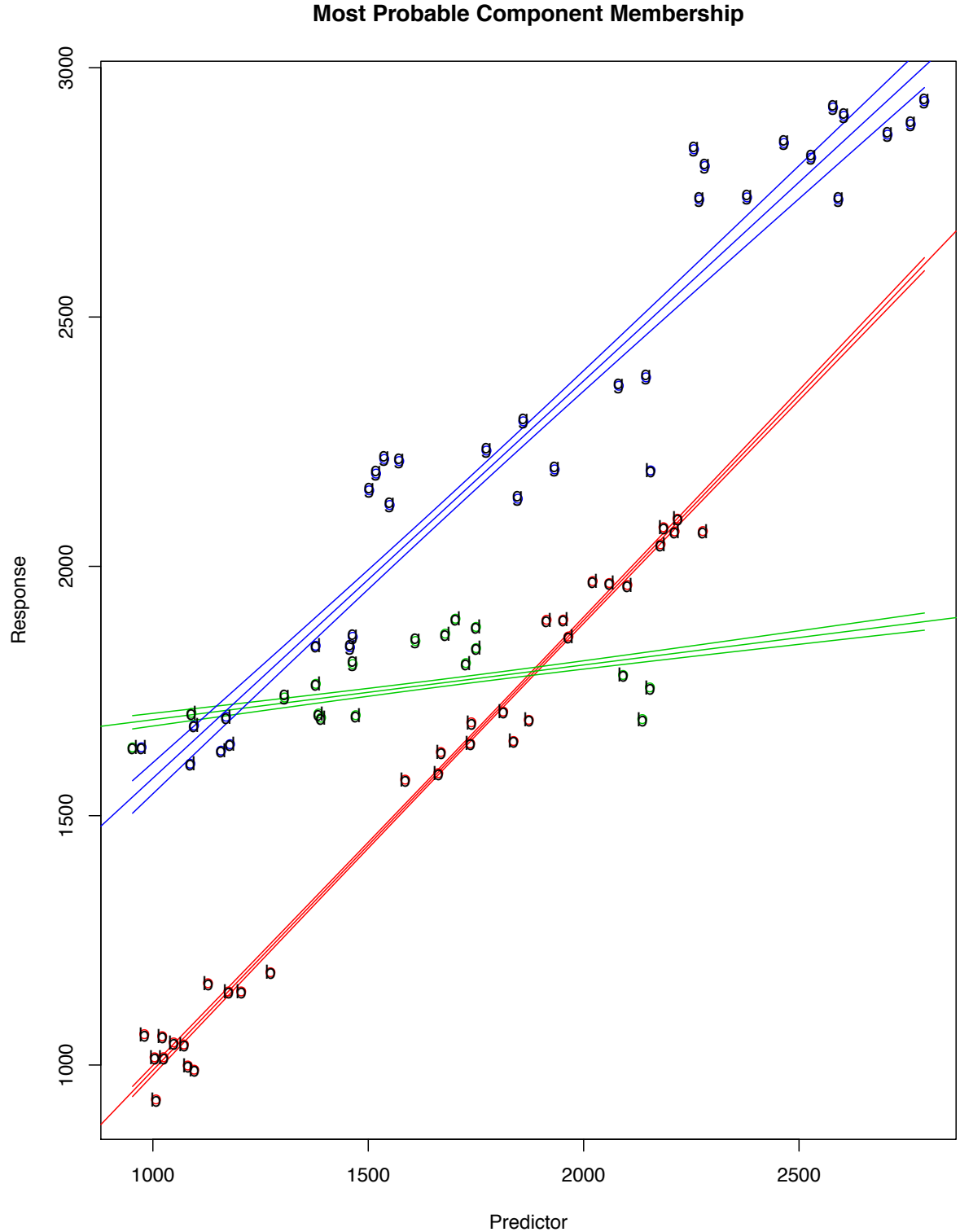
Mixture of Regressions (flexmix)

best model (AIC, BIC): *3 categories*





Mixture of Regressions (mixtools)



Summary

- Locus equations define a space (F2vowel x F2onset) and a class of models in that space (linear regressions)
- Here we can find the six phones with GMMs (but no voicing variation here)
- But we can also find the three phonemes with MGLMs
- Need to implement sampling for this

Conclusions

- Mixtures of regressions (or, GLM's) can find categories and relationships simultaneously
- Hard to find phones, and then hard to group into phonemes
- No direct inference of a surface inventory (how many kinds of /g/?),
- phones are epiphenomenal

Implications

- Phonotactic learning?
- Incomplete neutralization as T's to the same mean (but with different distributions)? No identification as "same phone".
- No allophonic feeding? (chain shifts)
- E-language vs. I-language?