A comparison of four vowel overlap metrics

Matthew C. Kelley

ACOL 2017
October 28, 2017

Alberta Phonetics Laboratory

UNIVERSITY OF ALBERTA
DEPARTMENT OF LINGUISTICS

Acknowledgments

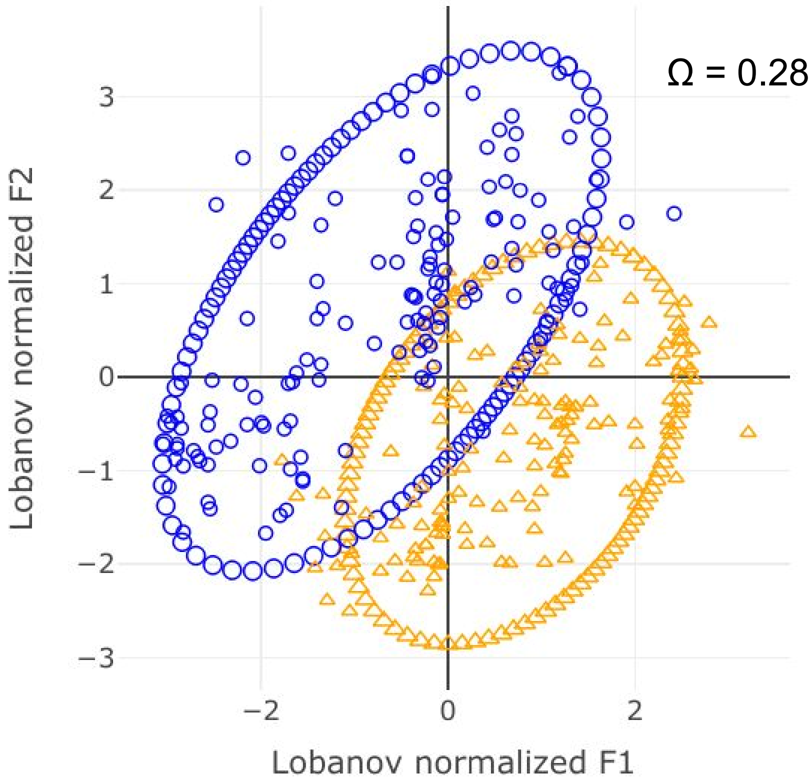
- Geoffrey Stewart Morrison (Aston University)
- Benjamin V. Tucker (University of Alberta)
- Alicia Wassink (University of Washington)

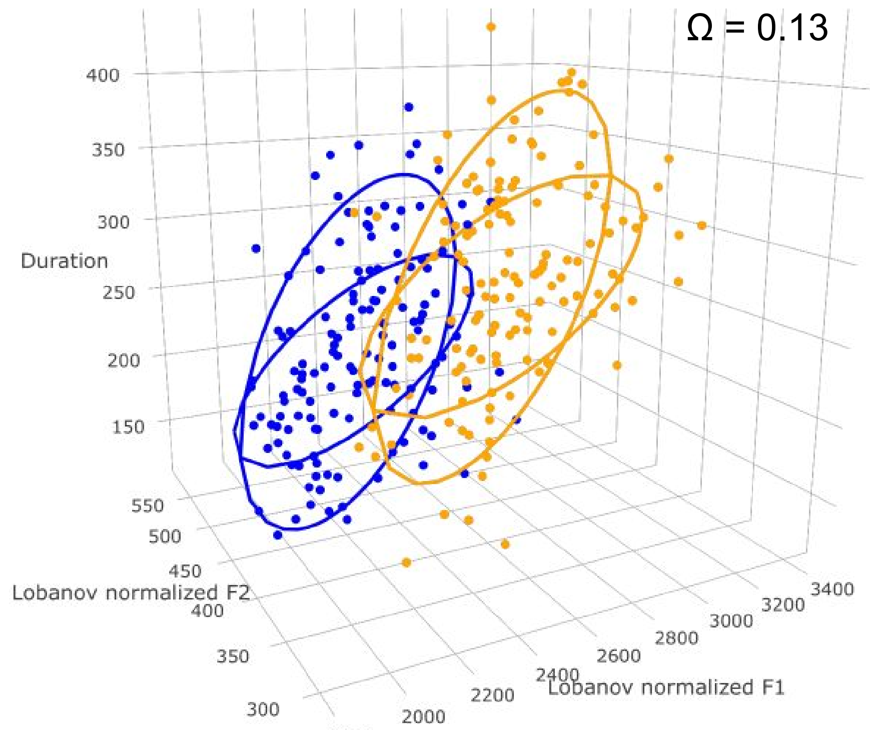
Introduction: Vowel Overlap

- Quantitative measures of vowel overlap useful in various fields
    - Sociophonetics
    - Dialectology
    - Second-language speech learning
- Ideally, such a metric would be
    - Accurate: gives desired results
    - Precise: gives similar results on similar data

Introduction: Spectral overlap assessment metric

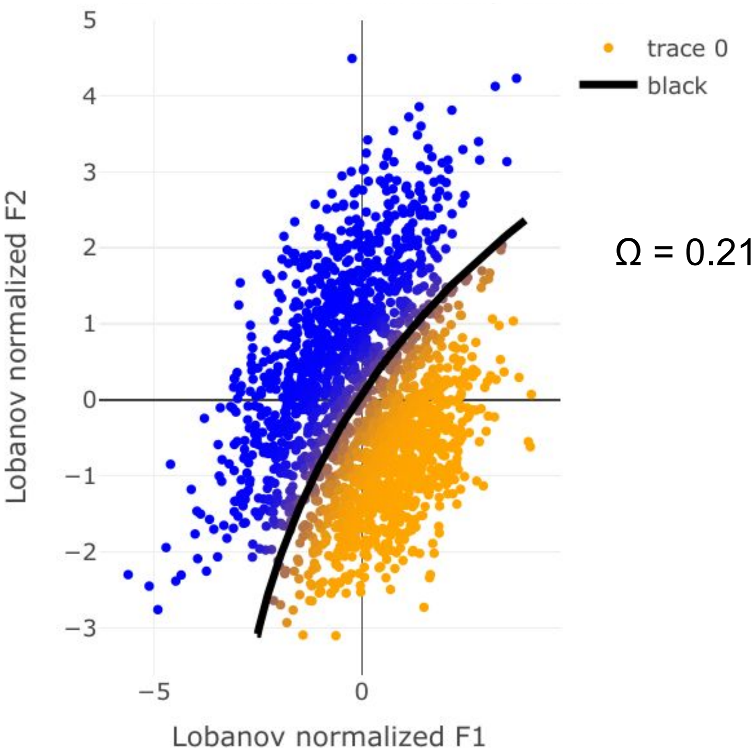
- Proposed by Wassink (2006)
- Represents data as ellipses or ellipsoids

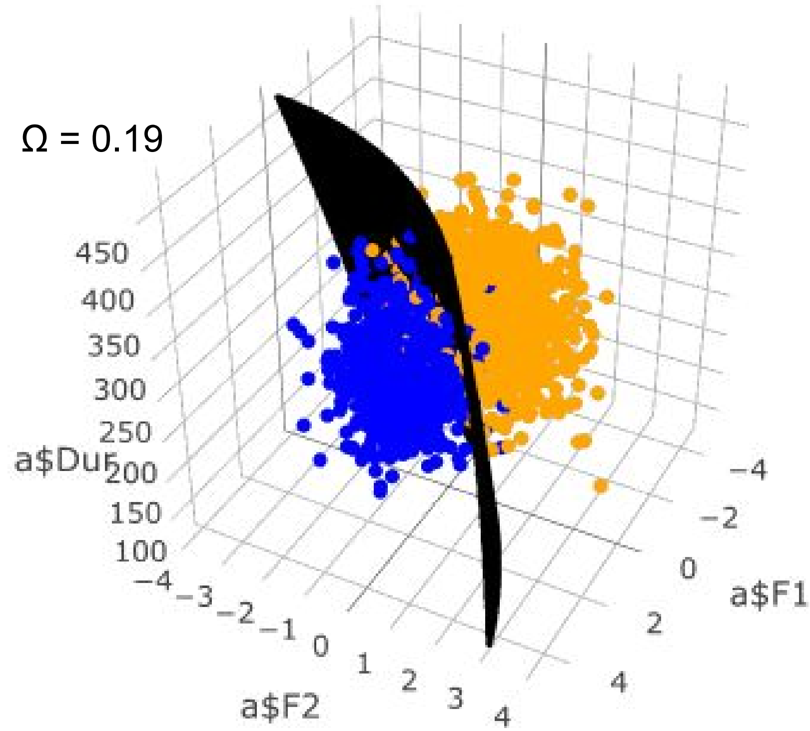




Introduction: *A posteriori* probability-based metric

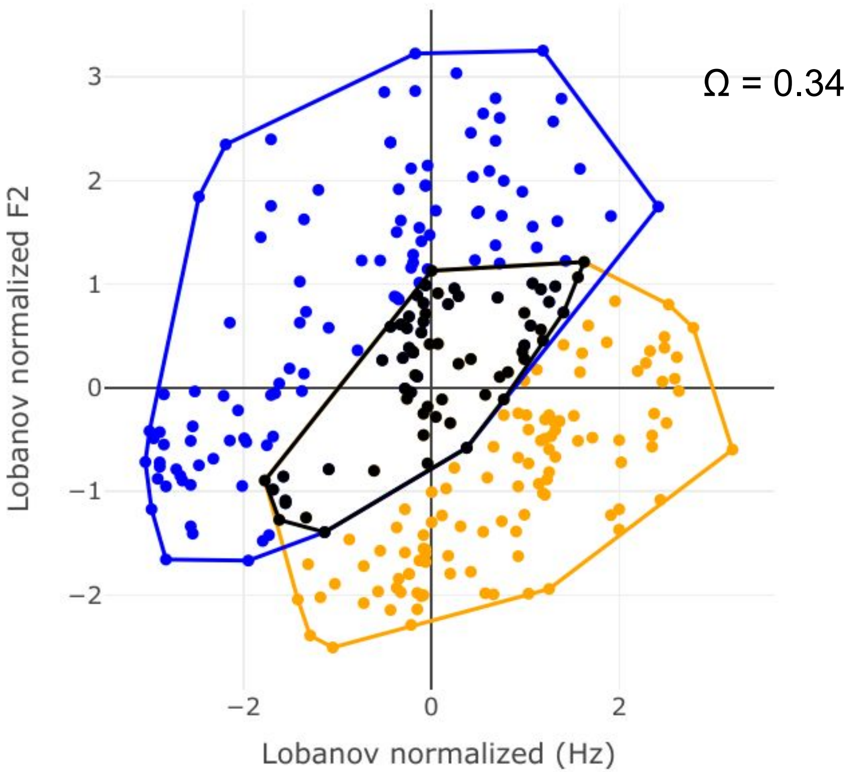
- Proposed by Morrison (2008)
- Quadratic discriminant analysis on data points
- Then uses posteriors to determine overlap

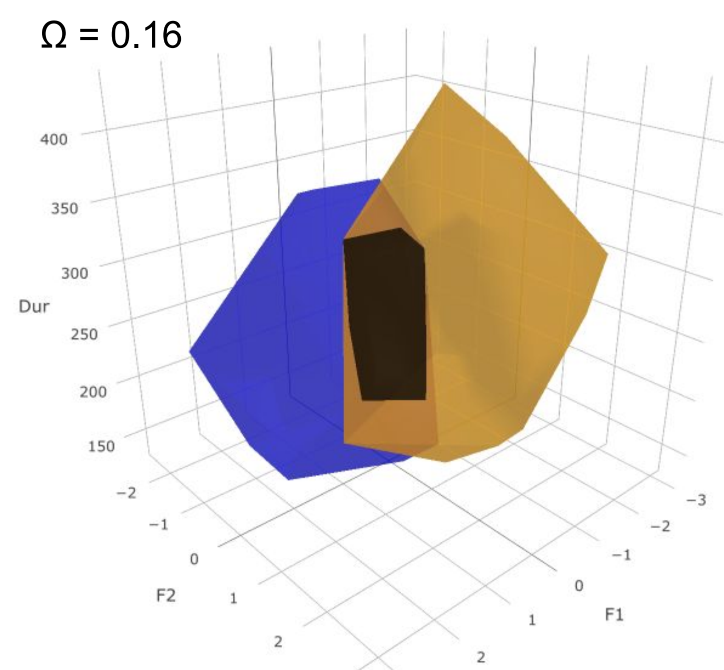




Introduction: Vowel overlap assessment with convex hulls metric

- Proposed by Haynes & Taylor (2014)
- Represents categories and overlap with convex hulls





Introduction: Pillai score

- Introduced by Hay et al. (2006)
- Uses Pillai score from running a MANOVA analysis
- Isn’t designed for visualization

Introduction: Research Question

- Nycz and Hall-Lew (2015) claim that the Spectral Overlap Assessment Metric is better for measuring overlap than Euclidean Distance, linear mixed-effects regression, and Pillai score
  - But what about *a posteriori* and convex hulls? How do they compare to SOAM and Pillai?
- In terms of accuracy and precision, **which proposed measure fares the best?**

Methodology

- Metrics implemented in R
- Monte Carlo simulation run to compare metrics along accuracy and precision
- Data generated randomly using mrnorm() from MASS (Ripley et al., 2016)
- Seeded with mean, std. dev., and covariance for all [i] and [ɪ] tokens from Hillenbrand et al. (1995), normalized with Lobanov technique from phonTools (Barreda, 2015)
- Use 1,000 generated samples for *a posteriori*

Methodology: Accuracy Simulations

- Calculate metrics 1000 times on three different kinds of data sets
  - 1. We want overlap of 0
    - Separate data points by pushing means apart and other points accordingly
  - 2. We want overlap of 0.5
    - Move categories on top of each other until Jenson-Shannon divergence was approximately 0.5

Methodology: Accuracy Simulations (cont)

- Calculate metrics 1000 times on three different kinds of data sets
  - 3. We want overlap of 1
    - Using same information as seeds for [i] and [ɪ]

Results: 2D accuracy (mean absolute error)

Measure	Mean for 0 overlap	Mean for 0.5 overlap	Mean for desired 1 overlap
SOAM	0	0.14	0.09
A posteriori	4e-62	0.09	0.07
Convex Hulls	0	0.12	0.29
Pillai	0.005	0.11	0.04

Results: 2D precision (mean absolute deviance from the median)

Measure	Mean for 0 overlap	Mean for 0.5 overlap	Mean for desired 1 overlap
SOAM	0	0.09	0.06
A posteriori	0	0.08	0.03
Convex Hulls	0	0.11	0.08
Pillai	0	0.07	0.02

Results: 3D accuracy (mean absolute error)

Measure	Mean for 0 overlap	Mean for 0.5 overlap	Mean for desired 1 overlap
SOAM	0	0.11	0.26
A posteriori	5e-73	0.08	0.13
Convex Hulls	0	0.33	0.65
Pillai	0.004	0.11	0.05

Results: 3D precision (mean absolute deviance from the median)

Measure	Mean for 0 overlap	Mean for 0.5 overlap	Mean for desired 1 overlap
SOAM	0	0.11	0.13
A posteriori	0	0.08	0.04
Convex Hulls	0	0.07	0.06
Pillai	0	0.07	0.03

Discussion: Accuracy

- *A posteriori* performed the best overall
  - Output a value that was effectively 0 in 0 overlap cases
  - Most accurate in middle range where it matters most
  - Didn’t perform the best when we wanted 1
    - Still consistently high values, so less important that it reaches 1

Discussion: Precision

- Pillai score performed the best
- *A posteriori* was competitive, though
- Spectral overlap assessment metric likely suffered in 3D due to the way the ellipsoid axes are calculated, since duration is on a different scale than normalized F1 and F2

Discussion: Recommendation

- Use the *a posteriori* probability-based metric
  - It performs the best in accuracy, and its near-best performance in precision make it the all-round winner
  - Additionally, since a classifier is learned, more things can be done with it

Conclusion: Future Research Avenues

- Can Pillai score be visualized meaningfully?
- 2D projections of 3D plots
- Explore ways to improve metrics
  - E.g., better figure fitting for spectral overlap assessment metric
- Explore ways to improve *a posteriori* plots
- Examine the measures' performance on real data sets

References

Barreda, S. (2015). phonTools: Functions for phonetics in R. (Version 0.2-2.1) [Software]. Available from https://cran.r-project.org/web/packages/phonTools/index.html

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.

Haynes, E. F., & Taylor, M. (2014). An assessment of acoustic contrast between long and short vowels using convex hulls. *The Journal of the Acoustical Society of America*, *136*(2), 883–891.

Hillenbrand, J. et al. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97(5), 3099-3111.

Morrison, G. S. (2008). Comment on "A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties" [J. Acoust. Soc. Am.119, 2334–2350 (2006)]. *The Journal of the Acoustical Society of America*, *123*(1), 37–40.

Nycz, J., & Hall-Lew, L. (2015). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*(1), 060008.

References

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D. (2016). Modern Applied Statistics with S (Version 7.3-45) [Software]. Available from https://cran.r-project.org/web/packages/MASS/index.html

Wassink, A. B. (2006). A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. *The Journal of the Acoustical Society of America*, *119*(4), 2334–2350.