Quantifying and Interpreting Vowel Formant Trajectory Information

Alicia Beckford Wassink, U of Washington
Chris Koops, U of New Mexico

Best Practices in Sociophonetics Workshop
NWAV 42, October 17, 2013 (Pittsburgh, PA)

Framing Questions

- Why do sociophoneticians need vowel trajectory information?
- How do I extract and represent formant trajectory information?
- How do I normalize formant trajectory information?
- How do I compare formant trajectories statistically?
  - Independent comparisons at discrete time points
  - Smoothing Spline ANOVA approaches

useful links

- Our workshop repository (http://faculty.washington.edu/wassink/NWAV2013)
  - these slides
  - soundfiles
  - datasets
  - R scripts
- R course website (http://faculty.washington.edu/wassink/LING580-RinLx/RinLxSyllabus.html)

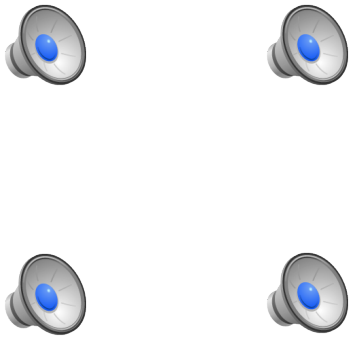
Listening Exercise

Seattle speakers' front vowels

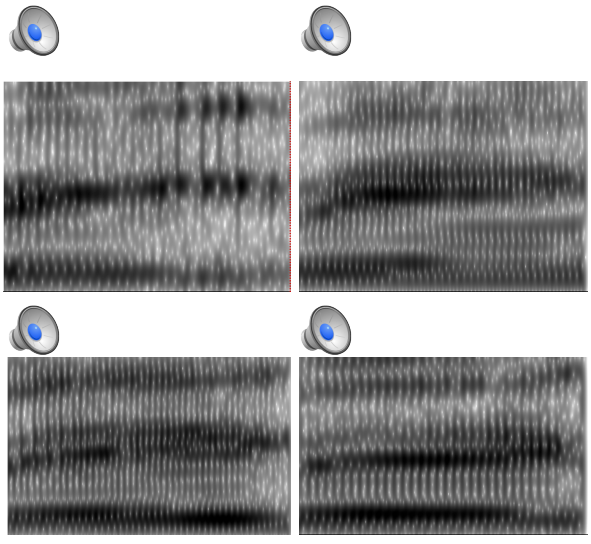
University of Washington, Department of Linguistics
National Science Foundation
The **Pacific Northwest English Study**

Research funded by National Science Foundation award BCS-1147678

Speaker 24 (female, EurAm)

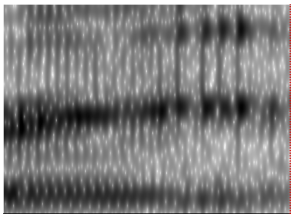


Speaker 24 (female, EurAm)

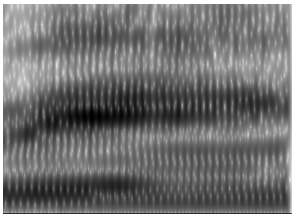


Speaker 24 (female, EurAm)

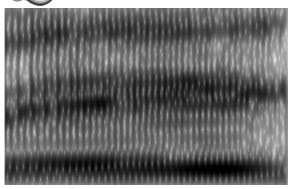
bait (ey)



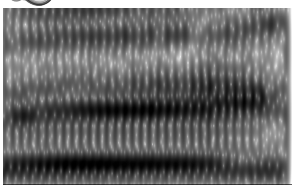
bang (æŋ)



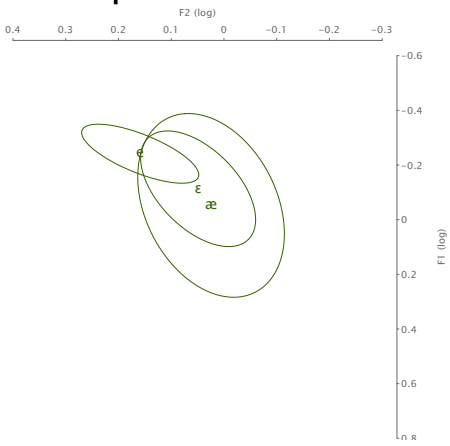
bag (æg)



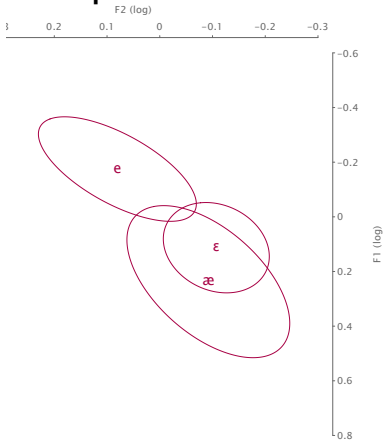
bɛg (ɛg)



Speaker 17

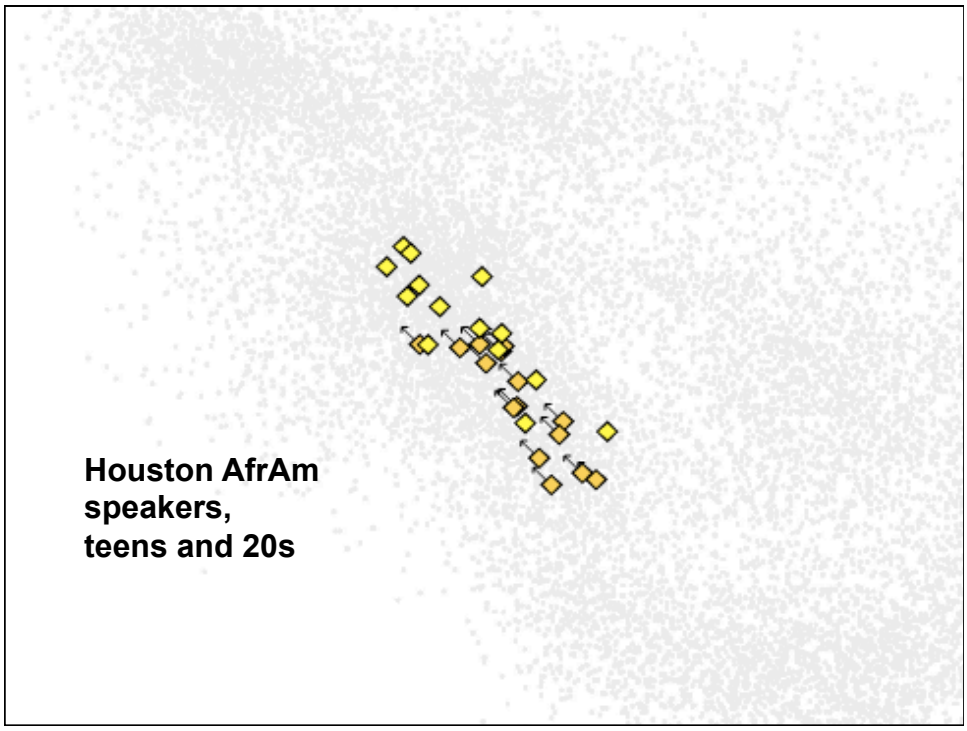


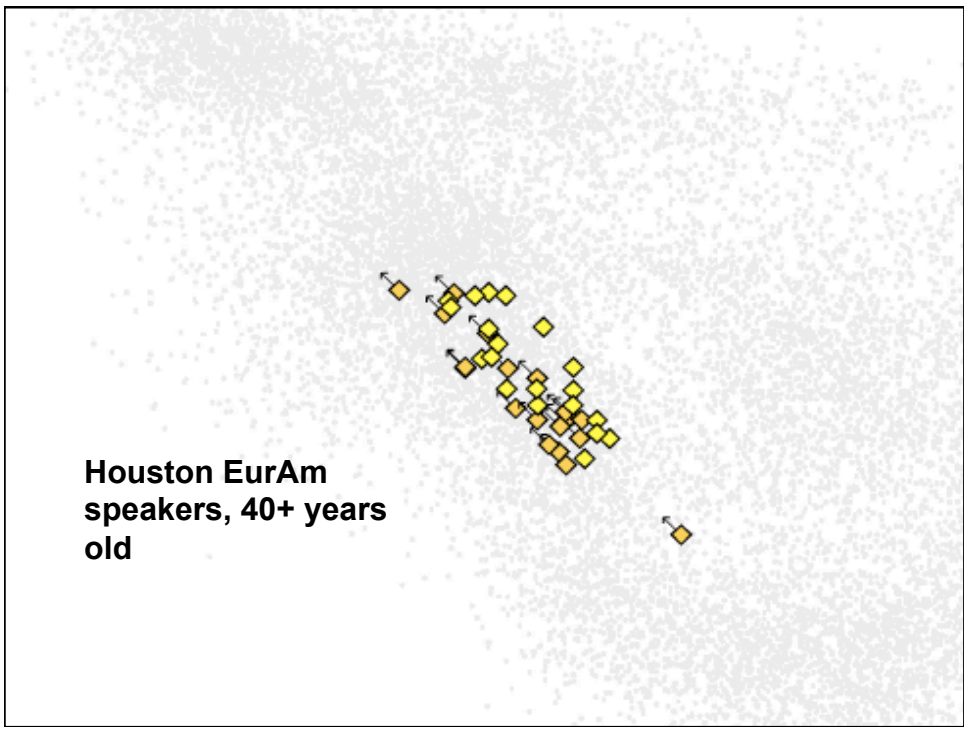
Speaker 24



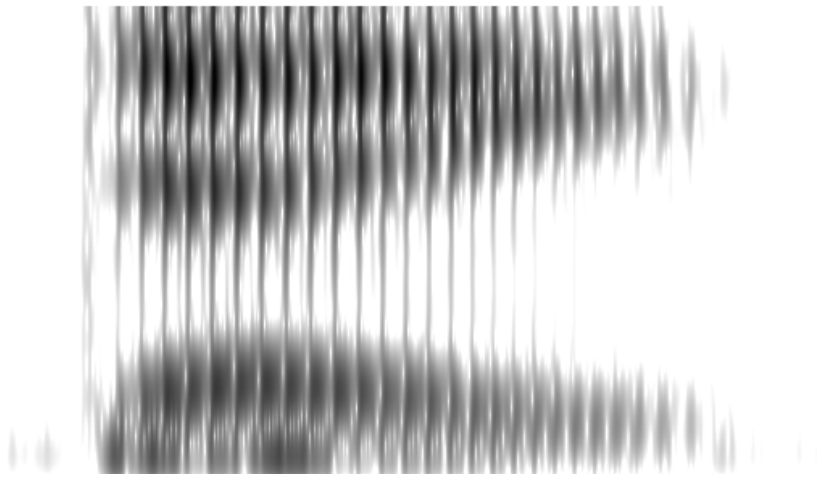
Listening Exercise

Houston Anglo and African American lowered /ey/

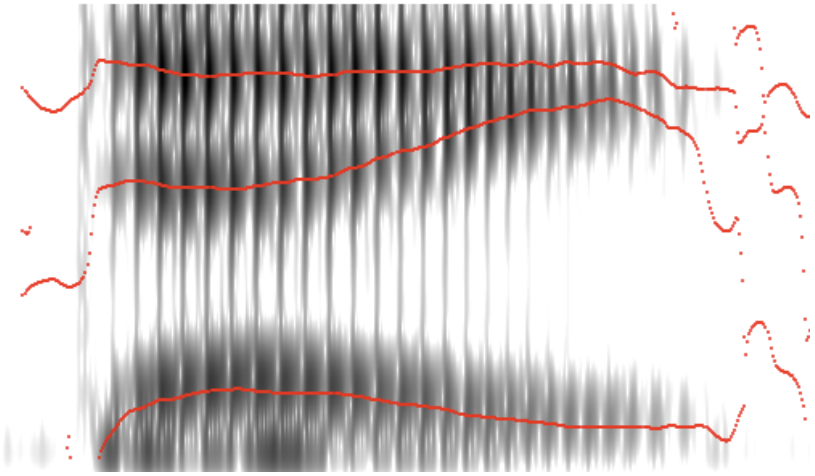




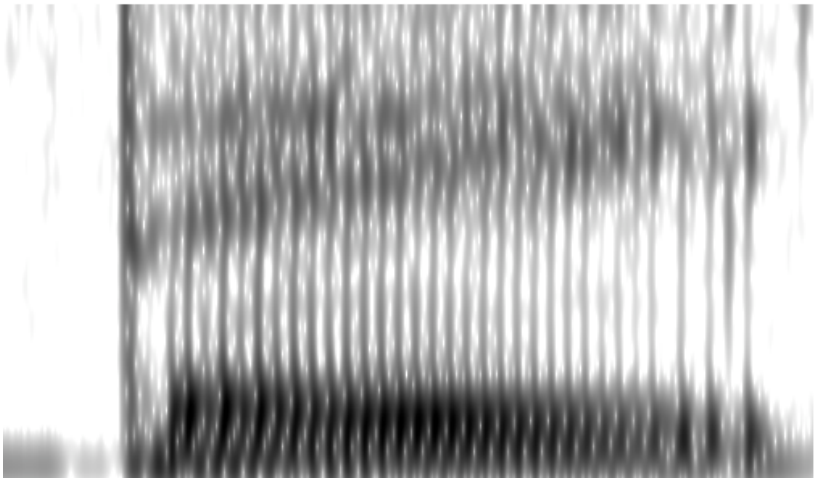
"date" (speaker JS, EurAm)



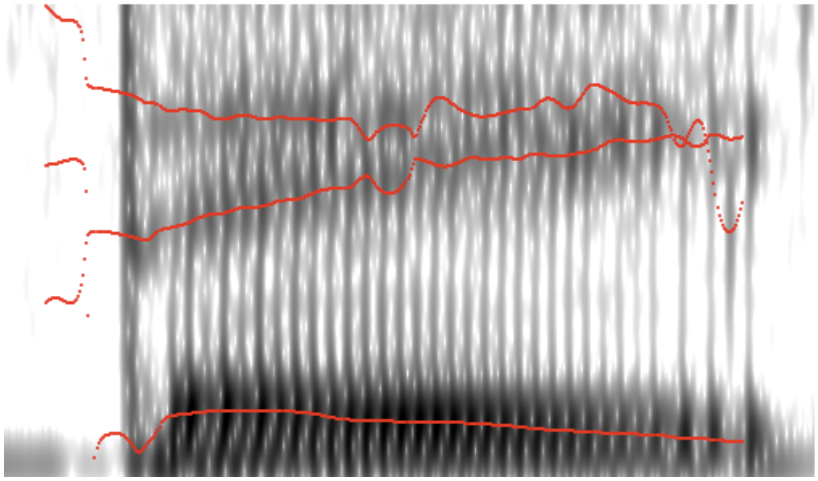
“date” (speaker JS, EurAm)



“date” (speaker BC, AfrAm)



"date" (speaker BC, AfrAm)



Vowel-Inherent Spectral Change

- definition: the relatively slowly varying changes in formant frequency associated with vowels (even in the absence of consonantal context). (Nearey and Assmann, 1986)
- abbreviated "VISC"

Framing Questions

- Why do sociophoneticians need vowel trajectory information?
- How do I extract and represent formant trajectory information?
- How do I normalize formant trajectory information?
- How do I compare formant trajectories statistically?
  - Independent comparisons at discrete time points
  - Smoothing Spline ANOVA approaches

Why study vowel trajectories? Reason 1:

- Auditory transcription can underrepresent critical variation (e.g., glide reduction, diphthongization or monophthongization strategies)
  - need more than nucleus and glide to describe what makes two sounds "sound different"

Why study vowel trajectories? Reason 2:

- (Even) acoustic analysis can underrepresent crucial variation if we don't represent the "right" information:
- Nearey and Assmann (1986): Need to distinguish 3 types of auditory/acoustic relations:
  - 1. traditional phonological or "nominal" monophthong (e.g., /ɪ, ɛ, æ, ʊ/)
  - 2. traditional phonological, phonemic or "true" diphthong (e.g. /aɪ, aʊ, ɔɪ/)
  - 3. "phonetic" diphthong (e.g., /eɪ/, /oʊ/, /ɛg/, /æg/)

Why study trajectories? Reason 3:

- Problem: how does phonetic structure relate to listeners' mental representations:
  - How much spectral change is required for a phone to be perceived as diphthongal? (Nearey & Assmann, 1986; Morrison, 2007; Jacewicz et al. 2011)
  - Duration is important! (Johnson, 2010)
  - Between dialect differences in timing and excursions of trajectories (Thomas, 2004)

Interim Summary

- Monophthongs don't always look monophthongal; nor diphthongs, diphthongal
- Dialects, or social groups may differ with regard to diphthongization strategy (e.g., Yaeger 1979; Heselwood, 2009)
- Can be tricky to measure phenomena (e.g., triphthongization, glide-weakening) when a boundary cannot be detected between "a" nucleus and "a" glide (Koops, 2010)

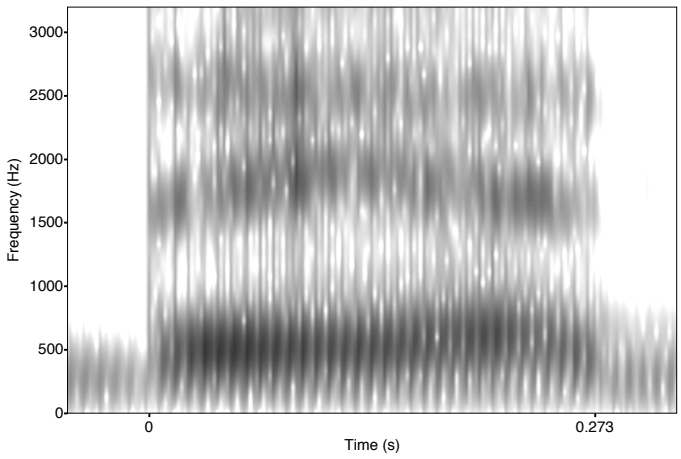
Framing Questions

- Why do sociophoneticians need vowel trajectory information?
- How do I extract and represent formant trajectory information?
- How do I normalize formant trajectory information?
- How do I compare formant trajectories statistically?
  - Independent comparisons at discrete time points
  - Smoothing Spline ANOVA approaches

Extracting trajectory information

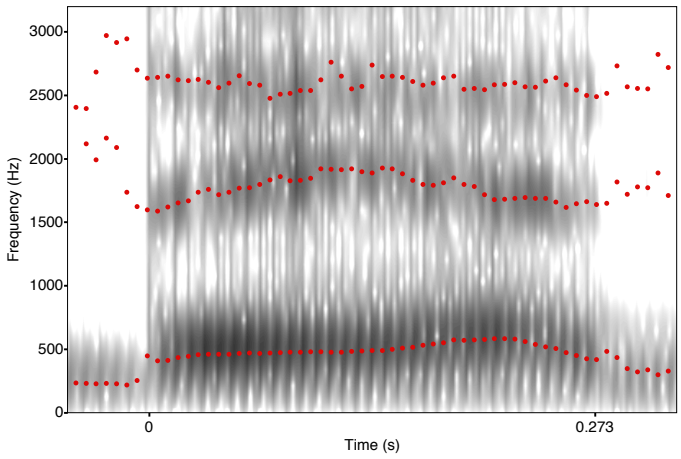
- Extracting a whole series of frequency values is not necessarily more involved than extracting single points
  - if done automatically, e.g. with a Praat script
  - and as long as the vowel boundaries are known
  - any number of intervals can be extracted
- Rather, one question is: How many sampling points should be collected? This depends on:
  1. how much temporal detail is needed
  2. how all of this detail will be stored
  3. how the information will be statistically evaluated

How much temporal detail is needed?



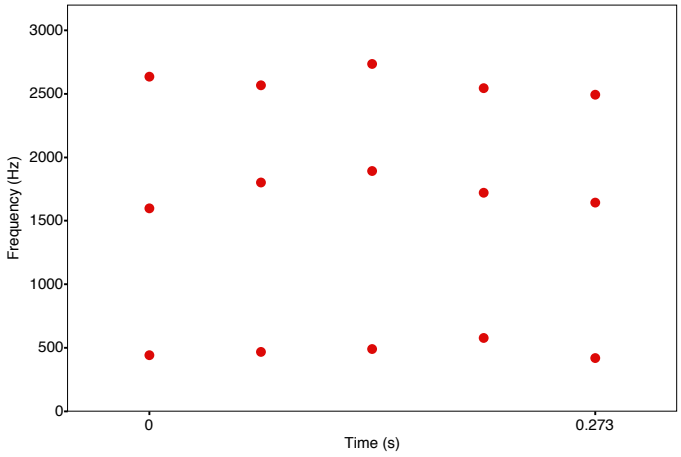
"dead"

How much temporal detail is needed?



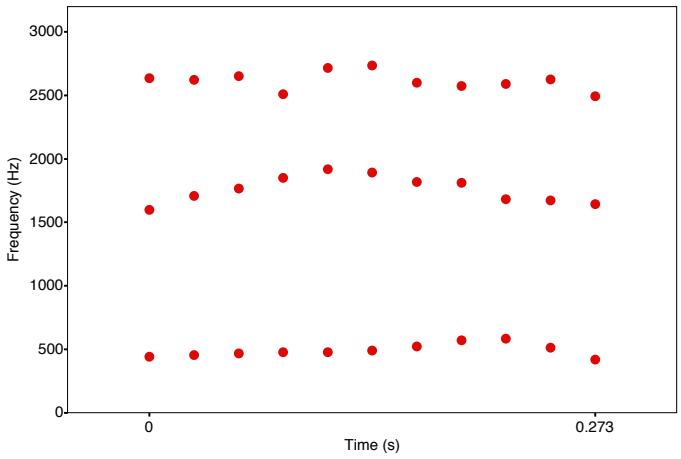
"dead"

How much temporal detail is needed?



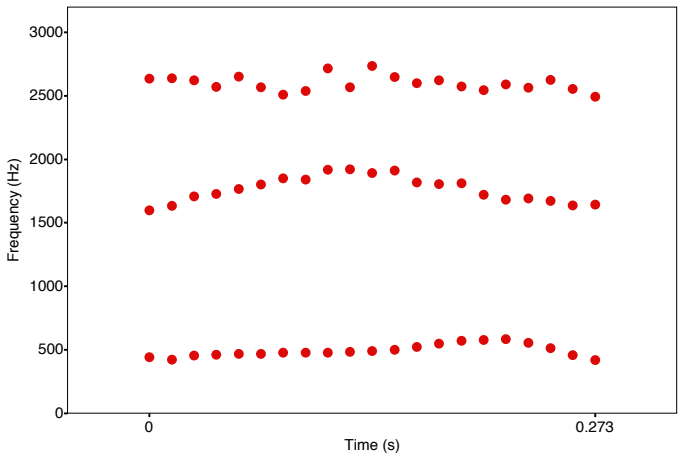
4 time steps

How much temporal detail is needed?



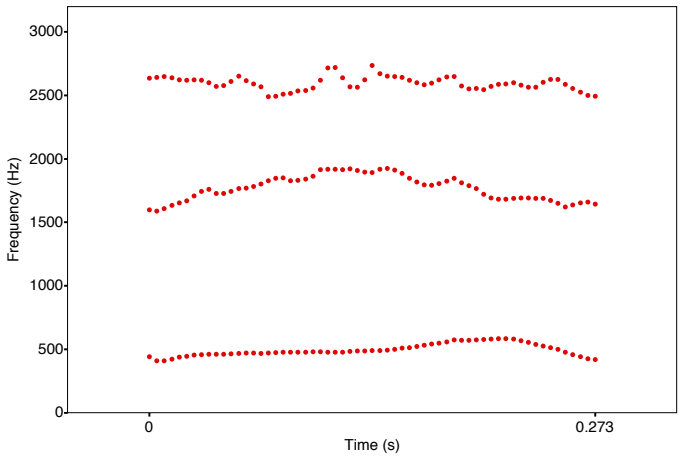
10 time steps

How much temporal detail is needed?



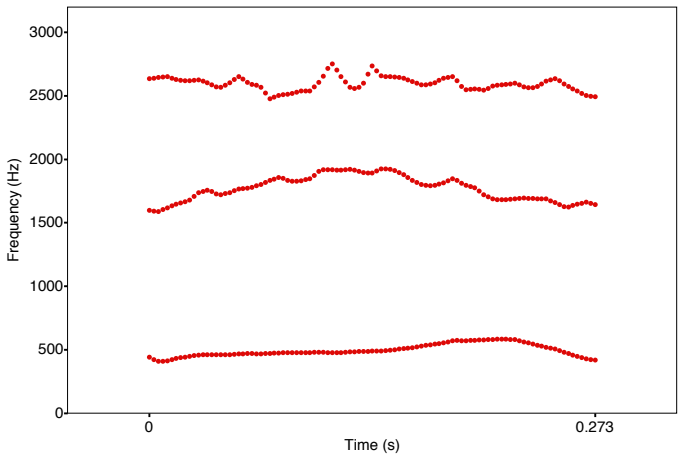
20 time steps

How much temporal detail is needed?



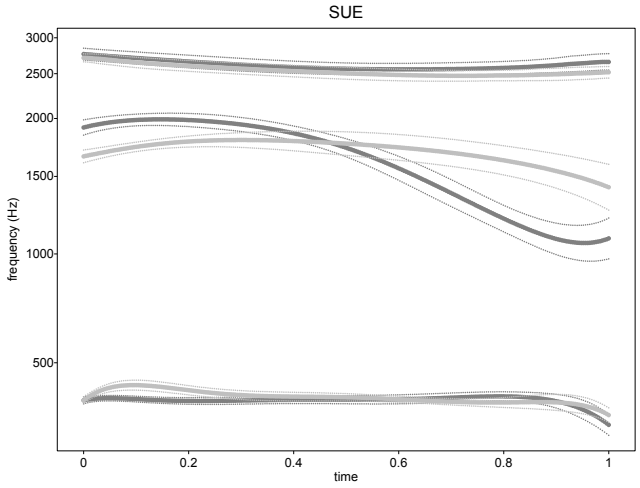
60 time steps

How much temporal detail is needed?



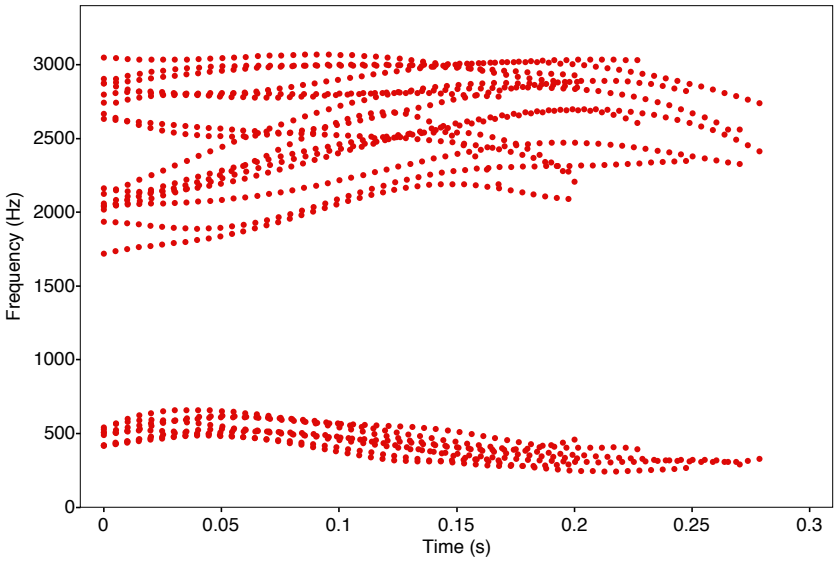
100 time steps

Example: two types of fronted /u/



(Koops 2010a)

The raw Houston /ey/-contours



How will all this detail be stored?

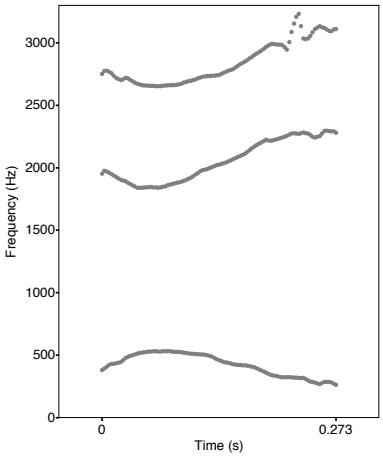
Grouping variables | Time points | Frequencies

	A	B	C	D	E	F	G	H
1	word	group	speaker	time	normTime	F1	F2	F3
2	DATE	EurAm	JS	0.0000	0.00	438	1991	2558
3	DATE	EurAm	JS	0.0182	0.05	458	1983	2787
4	DATE	EurAm	JS	0.0364	0.10	478	1976	2773
5	DATE	EurAm	JS	0.0545	0.15	498	1968	2760
6	DATE	EurAm	JS	0.0727	0.20	516	1960	2750
7	DATE	EurAm	JS	0.0909	0.25	533	1954	2741
8	DATE	EurAm	JS	0.1091	0.30	548	1948	2734
9	DATE	EurAm	JS	0.1273	0.35	562	1943	2728
10	DATE	EurAm	JS	0.1455	0.40	573	1939	2723
11	DATE	EurAm	JS	0.1636	0.45	582	1937	2719
12	DATE	EurAm	JS	0.1818	0.50	589	1936	2717

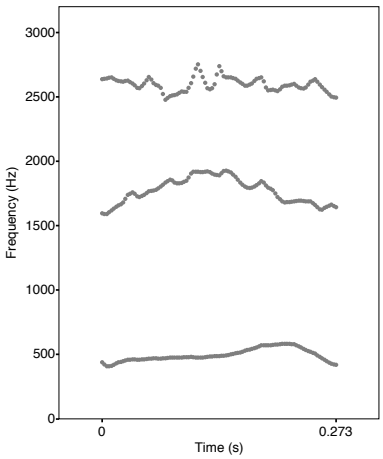
Modeling formant contours

- The extracted raw formant contours can be stored and further processed in simplified form by expressing their shape in the form of an appropriate mathematical function.
- Different mathematical functions approximate the shape of formants, especially diphthongs
  - **Cosine functions** (e.g. Watson & Harrington 1999, Harrington et al. 2008)
  - **Polynomials** (e.g. McDougall and Francis Nolan 2007, Morrison 2009, Risdal & Kohn, this conference)

Fitting polynomial functions

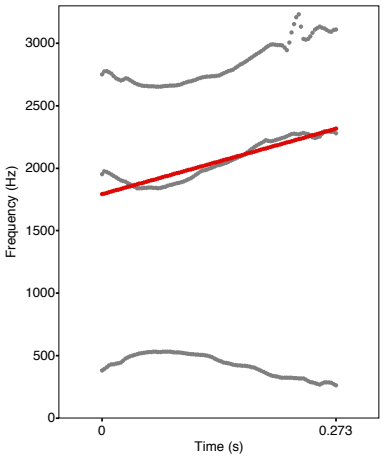


date

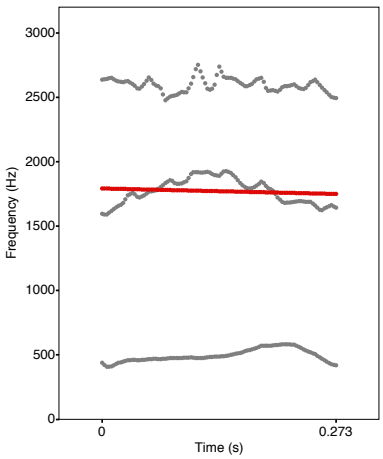


dead

1 coefficient: y = a + bx

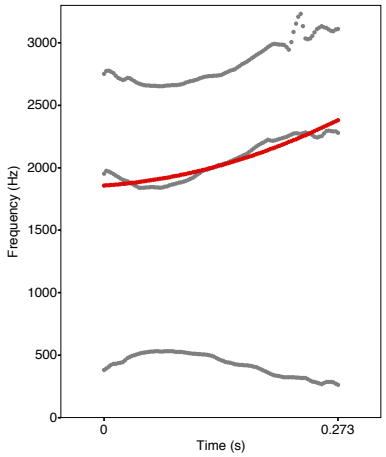


date

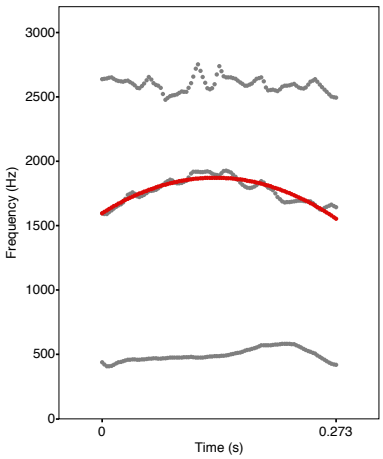


dead

2 coefficients: $y = a + bx + cx^2$

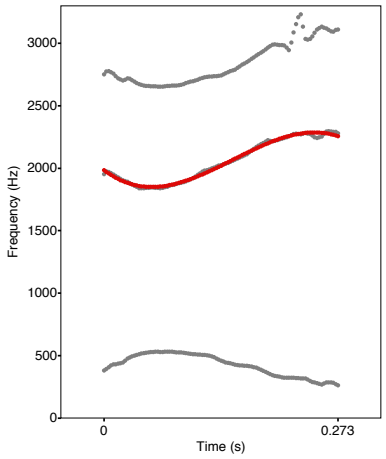


date

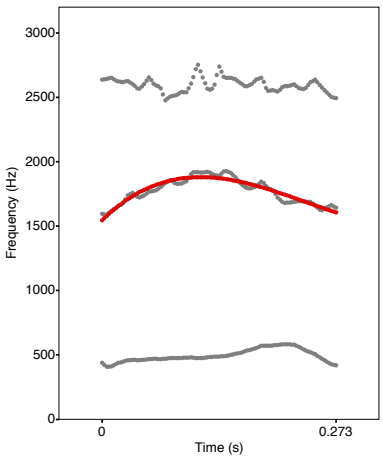


dead

3 coefficients: $y = a + bx + cx^2 + dx^3$

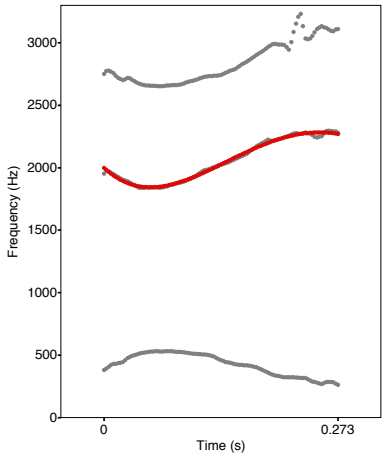


date

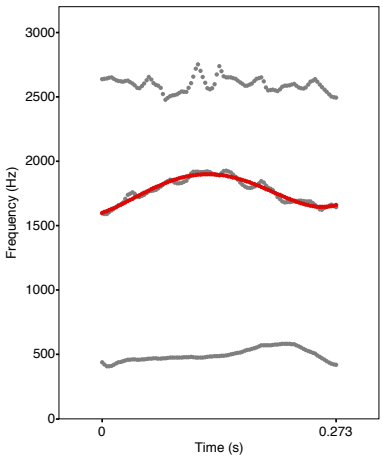


dead

4 coefficients: $y = a+bx+cx^2+dx^3+ex^4$



date



dead

Advantages of formant modeling

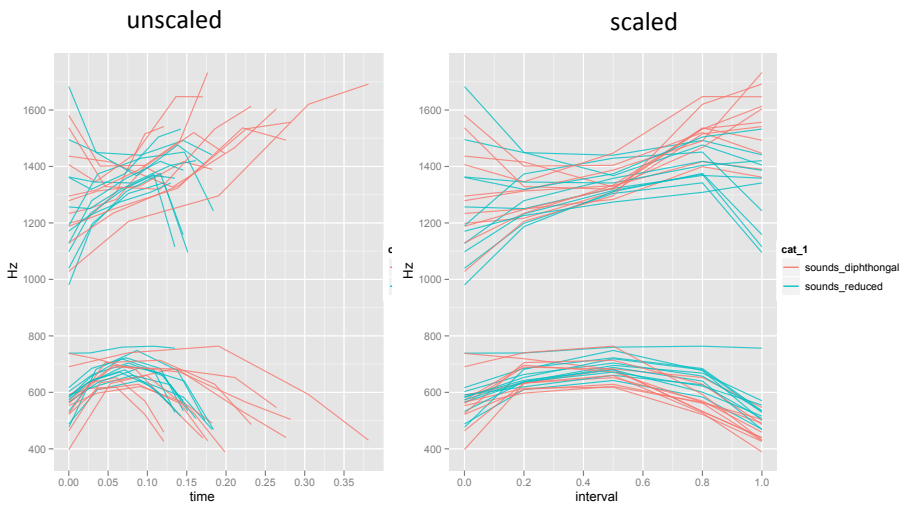
1. Less complex, more tractable representation
   - fewer values to be stored
   - in the case of some functions, one single coefficient can capture key aspects of the shape
     - e.g. cubic coefficients as index of relative degree of diphthongization (Risdal & Kohn, this conference)
- Smoothing
   - weakens the effect of signal noise and LPC errors
   - Same ideas as the 'smoothing' in Smoothing Spline ANOVA

Framing Questions

- Why do sociophoneticians need vowel trajectory information?
- How do I extract and represent formant trajectory information?
- How do I normalize formant trajectory information?
- How do I compare formant trajectories statistically?
  - Independent comparisons at discrete time points
  - Smoothing Spline ANOVA approaches

Why Normalize Trajectory Information

- Formant normalization and duration normalization (scaling) necessary
- …allows factoring out of physical (e.g., biological, sex-related) differences …
- …while preserving between-group, -dialect, or -language ones
- conversion to a perceptual scale enables us to reflect the perceptual primacy (Watt et al. 2011) of changes in F1 over changes in F2
- Formant normalization permits us to reflect a common extent in movement in the vowel spaces of different speakers
- Duration scaling allows comparison of information representing similar landmarks in different vowels



source: Michael Scanlon

How to Normalize

- *formant*: vowel-extrinsic, formant-intrinsic methods recommended, e.g.: Nearey 1 (logHz), Watt and Fabricius (s-centroid)
- Normalize point measurements within-speaker
- NORM, phonR, R, or Excel
- *duration*: proportional representation recommended: e.g., 20-50-80%, or ratio of `BeginTime-MyPointTime/SyllableDuration`, `BeginTime-MyPointTime/WordDuration`

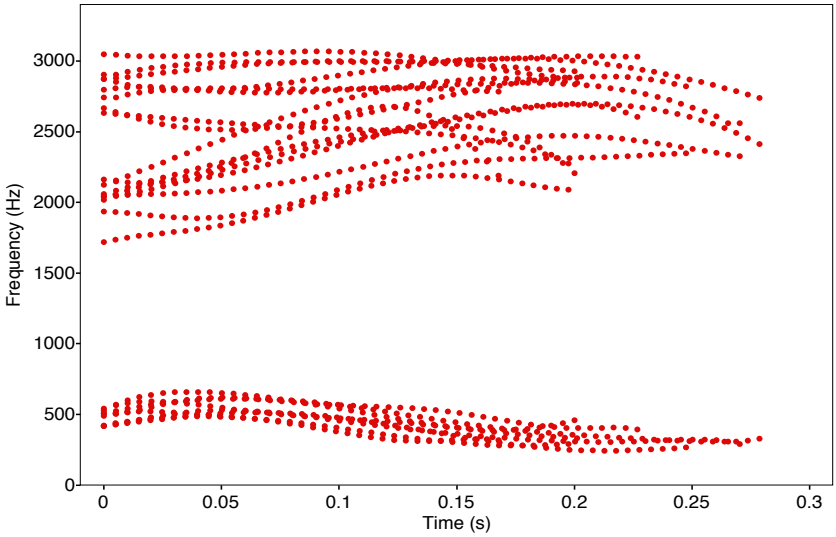
But…?

- Is it worth it to keep track of all these datapoints?
  - *yes*. comparing contours is less restrictive than traditional dichotomies such as diphthong vs. monophthong, nucleus vs. glide, steady state vs. non-steady states
  - no assumptions regarding which 1 or 2 points are "more representative" than others
- How do I usefully compare information from all of these timepoints?

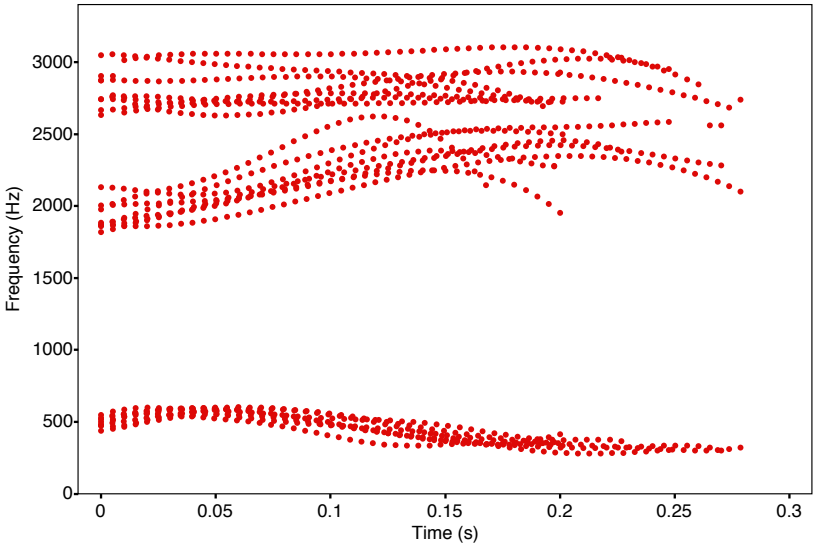
Framing Questions

- Why do sociophoneticians need vowel trajectory information?
- How do I extract and represent formant trajectory information?
- How do I normalize formant trajectory information?
- How do I compare formant trajectories statistically?
  - Independent comparisons at discrete time points
  - Smoothing Spline ANOVA approaches

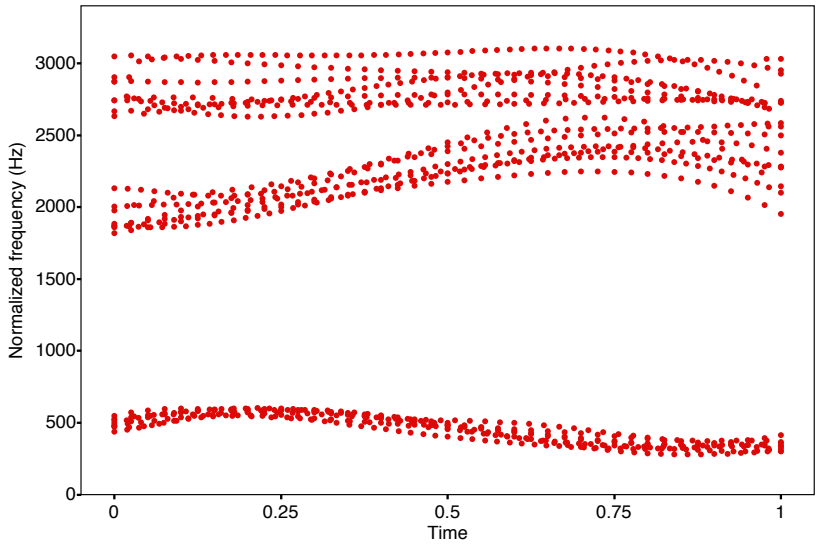
Houston /ey/-contours: raw



Nearey-1 normalized



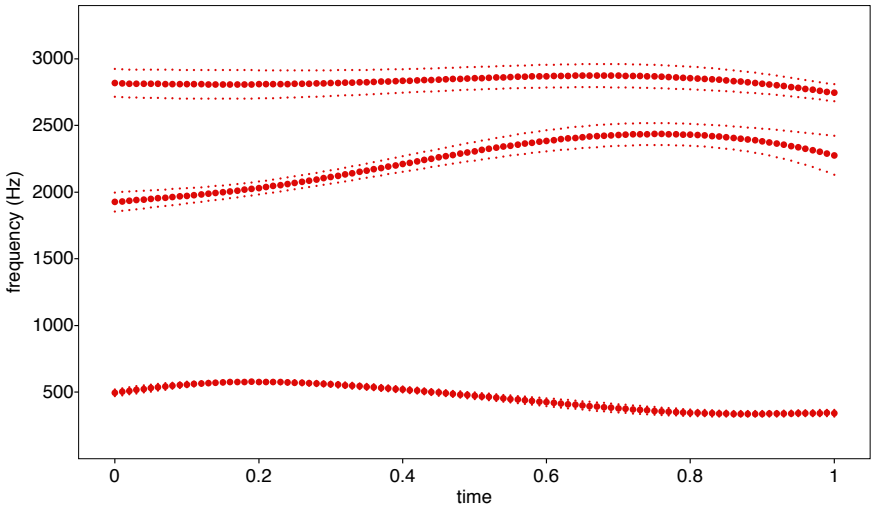
Time-scaled



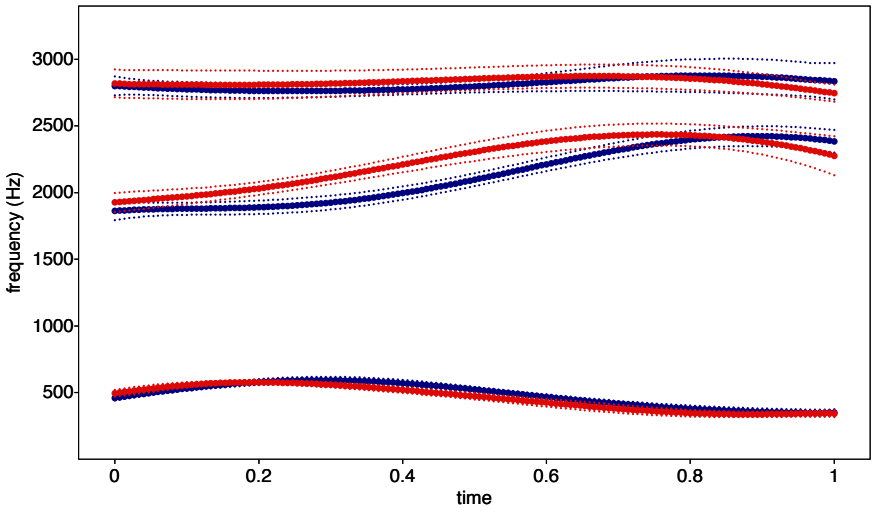
Statistical comparison: First, a more intuitive approach

- We can make a large number of individual statistical comparisons at discrete timepoints
  - For example, 100 time steps amount to a practically continuous contour where every point of potential relevance is covered
  - At each time step, then, a mean, standard error, and 95% confidence intervals can be calculated and plotted
  - whether two distributions overlap can then be determined from the resulting plot

Houston AAE lowered /ey/



Houston AAE and EAE lowered /ey/



Smoothing Spline ANOVA

- Designated statistical technique for the holistic comparison of curves (Gu 2002)
  - Used in linguistics by Davidson (2006) to analyze of ultrasound imaging data of tongue shapes
  - So far, only sporadically used for formants (e.g. Baker 2006, Nycz & DeDecker 2006, Koops 2010b)
  - `ssanova()` function included in several R packages
  - See the code in sample SSANOVA script and dataset for this workshop

Smoothing Spline ANOVA

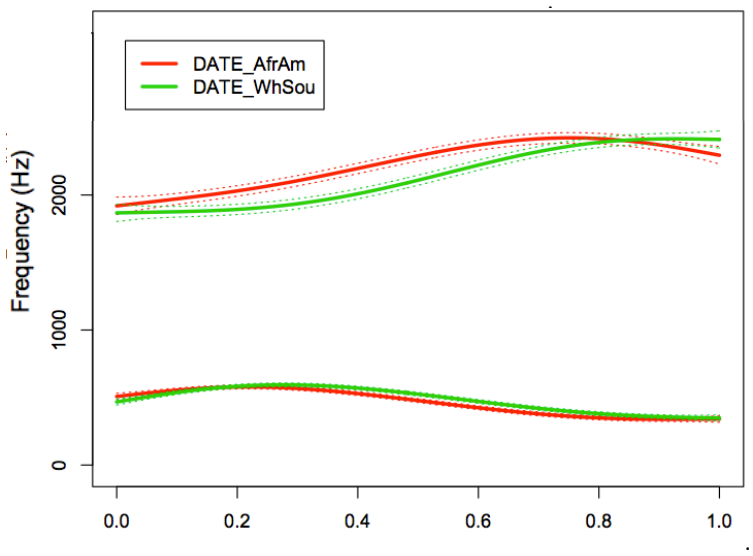
- Takes as input the time- /frequency- normalized frequency values of both groups to be compared
  - As in the sample data file shown earlier

1. Create model via function call, e.g.:

```
f2.model <- ssanova(F2 ~ Following.Phone + Interval +
                                Following.Phone:Interval,
                                data = aedata)
```

2. Print model summary with `summary()`
3. Obtain fitted model and standard error with `predict()`
4. Plot predicted values with confidence intervals
5. See sample code in file NWAV-demo-SSANOVA-RCode-new.txt

SS ANOVA plot



(Koops 2010b)

References

Baker, Adam. 2006. Quantifying diphthongs: a statistical technique for distinguishing formant contours. Paper presented at NWAV 35, The Ohio State University.

Baker, A., Mielke, J., & Archangeli, D. (2008). More Velar than /g/: Consonant Coarticulation as a Cause of Diphthongization. In C. B. Chang & H. J. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics* (60–68). Somerville, MA: Cascadilla Proceedings Project.

Benson, E. J., Fox, M. J., & Balkman, J. (2011). The Bag That Scott Bought: The Low Vowels in Northwest Wisconsin. *American Speech* 86 (3), 271-311.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. Journal of the Acoustical Society of America 120: 407-415.

diPaolo, M., Yaeger-Dror, M. & Wassink, A. B. (2011) Analyzing Vowels, ch. 8. In diPaolo, M. and Yaeger-Dror, eds. *Sociophonetics: a Student's Guide,* London: Routledge.

Foulkes, P., Scobbie, J. & Watt, D. (2010) Sociophonetics. In W. Hardcastle, J. Laver, and F. Gibbon, eds. *Handbook of Phonetic Sciences, 2nd ed. Oxford: Blackwell, 703-54*

Fruehwald, J. (2010) SS ANOVA. accessed online:

Gu, Chong. (2002) *Smoothing Spline ANOVA Models*. Springer, New York

Harrington, Jonathan, Felicitas Kleber and Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. JASA 123: 2825-2835

Heselwood, B. (2009) "Rhoticity without F3: low pass filtering, F1-F2 relations, and the perception of rhoticity," *Leeds Working Papers in Linguistics and Phonetics,* 52:49-64

Jacewicz, E. Fox, R. and Salmons, J. (2011) Cross-generational vowel change in American English, *Language Variation and Change*, 23 (2011), 45–86.

Johnson (2010) Speaker normalization in speech perception. accessed online: 10/16/2013 (http://linguistics.berkeley.edu/~kjohnson/papers/revised_chapter.pdf)

Koops, Chris. (2010a). /u/-fronting is not monolithic: Two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers* 16(2).

----- (2010b). Southern /e/ and /ɛ/ variants in Houston EAE and AAE: An analysis of formant contours. Paper presented at ADS, Baltimore.

Labov, W. (1994) *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

----- (2001) *Principles of Linguistic Change: Social Factors.* Oxford: Blackwell.

Labov, W., Ash, S. and Boberg, C. (2006) *The Phonological Atlas of North American English*. Berlin: Mouton de Gruyter.

McDougall, Kirsty and Francis Nolan. 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. Proceedings of ICPhS XVI, 1825–1828.

References 2

Morrison, G. (2007) "Theories of vowel inherent spectral change: A review," unpublished ms. accessed 7/16/2011 from: http://geoff-morrison.net/

----- (2009) Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. JASA 125: 2387–97.

Nearey, T. & Assmann, P. F. (1986) Modeling the role of vowel inherent spectral change in vowel identification, *JASA,* 80: 1297-308.

Nycz, Jennifer & De Decker, Paul. (2006). A new way of analyzing vowels: Comparing formant contours using Smoothing Spline ANOVA. Poster presented at NWAV 35, The Ohio State University.

Purnell, T. C. (2008). Prevelar raising and Phonetic Conditioning: The Role of Labial and Anterior Tongue Gestures. *American Speech* 83.4, 373-402.

Reed, C. (1952). The pronunciation of English in the state of Washington. *American Speech*, 27(3): 186-189.

----- (1961). The pronunciation of English in the Pacific Northwest. *Language*, 37(4): 559-564.

Thomas, E. R. (2004) "Instrumental Phonetics" in J. K. Chambers, P. Trudgill and N. Schilling-Estes (eds.) *The Handbook of Lang. Variation and Change,* Oxford: Blackwell, 168-200.

Wassink, A. B., Squizzero, R., Scanlon, M., Schirra, R., & Conn, J. (2009). Effects of Style and Gender on Fronting and Raising of /æ/, /e:/ and /ɛ/ before /g/ in Seattle English. Paper presented at NWAV38 in Ottawa, ON, Canada.

Watson, Catherine and Jonathan Harrington. 1999. "Acoustic evidence for dynamic formant trajectories in Australian English vowels." JASA 106: 458-468.

Watt, D., A. Fabricius &T. Kendall (2011) More on Vowels: Plotting and Normalization, ch. 9. In diPaolo, M. and Yaeger-Dror, eds. *Sociophonetics: a Student's Guide,* London: Routledge.

Yaeger, M. (1979) "Context Determined Variation in Montreal French Vowels" PhD Dissertation, U of Pennsylvania

Zeller, C. (1997) The investigation of a sound change in progress: /ae/ to /e/ in Midwestern American English. Journal of English Linguistics, 25: 142-155.