Quan+fying and Interpre+ng Vowel Formant Trajectory Informa+on

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

Research funded by National Science Foundation award BCS-1147678
Speaker 24 (female, EurAm)
Speaker 24 (female, EurAm)

bait (ey)  

bang (æŋ)

bag (æg)  

bɛɡ (ɛɡ)
Listening Exercise

Houston Anglo and African American lowered /ey/

Houston AfrAm speakers, teens and 20s
Houston EurAm speakers, 40+ years old

“date” (speaker JS, EurAm)
“date” (speaker JS, EurAm)

“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. /ai, au, oy/)
  - 3. “phonetic” diphthong (e.g., /ei/, /ou/, /eɡ/, /æɡ/)

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

20 time steps
How much temporal detail is needed?

60 time steps

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: word, group, speaker

Time points:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>word</td>
<td>group</td>
<td>speaker</td>
<td>time</td>
<td>normTime</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>2</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0000</td>
<td>438</td>
<td>1991</td>
<td>2558</td>
</tr>
<tr>
<td>3</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0182</td>
<td>458</td>
<td>1983</td>
<td>2787</td>
</tr>
<tr>
<td>4</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0364</td>
<td>478</td>
<td>1976</td>
<td>2773</td>
</tr>
<tr>
<td>5</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0545</td>
<td>498</td>
<td>1968</td>
<td>2760</td>
</tr>
<tr>
<td>6</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0727</td>
<td>516</td>
<td>1960</td>
<td>2750</td>
</tr>
<tr>
<td>7</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.0909</td>
<td>533</td>
<td>1954</td>
<td>2741</td>
</tr>
<tr>
<td>8</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.1091</td>
<td>548</td>
<td>1948</td>
<td>2734</td>
</tr>
<tr>
<td>9</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.1273</td>
<td>562</td>
<td>1943</td>
<td>2728</td>
</tr>
<tr>
<td>10</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.1455</td>
<td>573</td>
<td>1939</td>
<td>2723</td>
</tr>
<tr>
<td>11</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.1636</td>
<td>582</td>
<td>1937</td>
<td>2719</td>
</tr>
<tr>
<td>12</td>
<td>DATE</td>
<td>EurAm</td>
<td>JS</td>
<td>0.1818</td>
<td>589</td>
<td>1936</td>
<td>2717</td>
</tr>
</tbody>
</table>

Frequencies: F1, F2, F3

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

1 coefficient: \( y = a + bx \)

**date**

**dead**
2 coefficients: \( y = a + bx + cx^2 \)

3 coefficients: \( y = a + bx + cx^2 + dx^3 \)
4 coefficients: \( y = a + bx + cx^2 + dx^3 + ex^4 \)

**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
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 \( \frac{\text{BeginTime-MypoPointTime}}{\text{SyllableDuration}} \), \( \frac{\text{BeginTime-MypoPointTime}}{\text{WordDuration}} \)
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
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
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

References


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


## References 2

http://geoff-morrison.net/  


