More on Normalization: What did that algorithm do to my data? And what do I see now?

Workshop 2a, NWAV 38,
Ottawa, 22nd October 2009

Anne Fabricius,
Roskilde University Denmark

UNIVERSITAS ROSKILDENSIS · IN TRANQUILLO MORS · IN FLUCTU VITA

Presentation topics

- Normalization: why and how?
  - Parameters of normalization
- What did that algorithm do?
  - Some procedures for comparing normalization outputs
- What do I see?
  - quantifying comparative parameters on vowel plots
- Discussion

Part 1: Normalization: why and how

- **factoring out** of physical (i.e. acoustic) differences
- Listeners can unconsciously compensate for absolute formant frequency differences for any given vowel phoneme
- cognitive process of normalization

More on why….

- The major motivation for normalizing data is that through normalized data one can directly and *quantitatively* compare speakers' and speaker groups' vowel productions with one another
  - 'in community X, for every decade we add to a speaker's year of birth, vowel Y is Z units fronter',
  - Or, the gradient of a line between vowel x's and vowel y's average locations has changed by n degrees over m generations

- Thus, standard practice in many types of acoustic phonetic studies

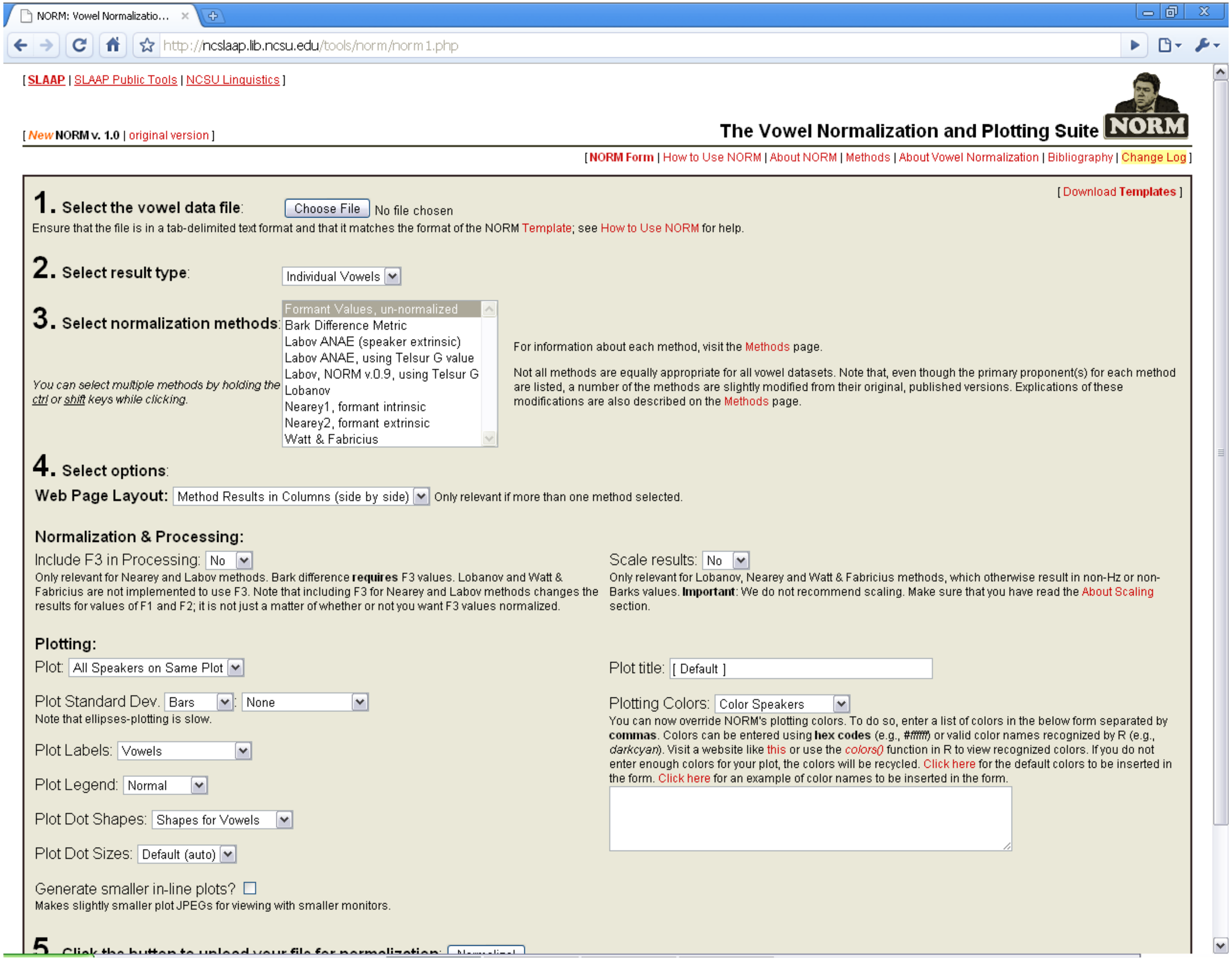
- Cross-gender and cross-age comparisons especially
- Same-gender comparisons
- Same speaker over time (age-dependent)

- Hence an obvious parameter for sociophonetic community studies

How: Two types of transforms...

- Mathematical transformations that are used on acoustic data are of two main types.

- The two types reflect the stages of processing carried out on an incoming acoustic signal
  - the peripheral auditory system (type 1)
  - the auditory processing centers of the brain (type 2)



NORM: Vowel Normalizatio...

http://ncslaap.lib.ncsu.edu/tools/norm/norm1.php

[SLAAP | SLAAP Public Tools | NCSU Linguistics]

[*New* NORM v. 1.0 | original version]

The Vowel Normalization and Plotting Suite NORM

[NORM Form | How to Use NORM | About NORM | Methods | About Vowel Normalization | Bibliography | Change Log]

[Download Templates]

1. Select the vowel data file: Choose File No file chosen

Ensure that the file is in a tab-delimited text format and that it matches the format of the NORM Template; see How to Use NORM for help.

2. Select result type: Individual Vowels

3. Select normalization methods:

- Formant Values, un-normalized
- Bark Difference Metric
- Labov ANAE (speaker extrinsic)
- Labov ANAE, using Telsur G value
- Labov, NORM v.0.9, using Telsur G
- Lobanov
- Nearey1, formant intrinsic
- Nearey2, formant extrinsic
- Watt & Fabricius

You can select multiple methods by holding the ctrl or shift keys while clicking.

For information about each method, visit the Methods page.

Not all methods are equally appropriate for all vowel datasets. Note that, even though the primary proponent(s) for each method are listed, a number of the methods are slightly modified from their original, published versions. Explications of these modifications are also described on the Methods page.

4. Select options:

Web Page Layout: Method Results in Columns (side by side) Only relevant if more than one method selected.

Normalization & Processing:

Include F3 in Processing: No

Only relevant for Nearey and Labov methods. Bark difference **requires** F3 values. Lobanov and Watt & Fabricius are not implemented to use F3. Note that including F3 for Nearey and Labov methods changes the results for values of F1 and F2; it is not just a matter of whether or not you want F3 values normalized.

Scale results: No

Only relevant for Lobanov, Nearey and Watt & Fabricius methods, which otherwise result in non-Hz or non-Barks values. **Important**: We do not recommend scaling. Make sure that you have read the About Scaling section.

Plotting:

Plot: All Speakers on Same Plot

Plot title: [Default]

Plot Standard Dev.: Bars : None

Note that ellipses-plotting is slow.

Plotting Colors: Color Speakers

You can now override NORM's plotting colors. To do so, enter a list of colors in the below form separated by **commas**. Colors can be entered using **hex codes** (e.g., #ffffff) or valid color names recognized by R (e.g., *darkcyan*). Visit a website like this or use the *colors()* function in R to view recognized colors. If you do not enter enough colors for your plot, the colors will be recycled. Click here for the default colors to be inserted in the form. Click here for an example of color names to be inserted in the form.

Plot Labels: Vowels

Plot Legend: Normal

Plot Dot Shapes: Shapes for Vowels

Plot Dot Sizes: Default (auto)

Generate smaller in-line plots? ☐

Makes slightly smaller plot JPEGs for viewing with smaller monitors.

5. Click the button to upload your file for normalization: Normalize!

Descriptive Parameters:

- 3-way cross-cutting parameters referring to where the algorithm derives its information from
  - Speaker extrinsic vs Speaker intrinsic
  - Vowel extrinsic vs Vowel intrinsic
  - Formant extrinsic vs Formant intrinsic

Scale/Method	**Speaker**	**Vowel**	**Formant**	**Reference; *NORM suite name***
Bark	intrinsic	intrinsic	intrinsic	Traunmüller 1990, 1997
mel	intrinsic	intrinsic	intrinsic	Stevens and Volkman 1940
Koenig	intrinsic	intrinsic	intrinsic	Koenig 1949
ERB	intrinsic	intrinsic	intrinsic	Moore and Glasberg 1983
Syrdal and Gopal	intrinsic	intrinsic	extrinsic	Syrdal and Gopal 1986
Bark Difference Metric	intrinsic	intrinsic	extrinsic	*Bark Difference Metric*
Gerstman	intrinsic	extrinsic	intrinsic	Gerstman 1968
Lobanov	intrinsic	extrinsic	intrinsic	*Lobanov*
Nearey $CLIH_{i4}$	intrinsic	extrinsic	intrinsic	*Nearey1*
Watt and Fabricius	intrinsic	extrinsic	intrinsic	Watt and Fabricius 2002; *Watt and Fabricius*
Nordström and Lindblom	intrinsic	extrinsic	extrinsic	Nordström and Lindblom 1975
Nearey $CLIH_{s4}$	intrinsic	extrinsic	extrinsic	Nearey 1977/8; *Nearey2*
Labov ANAE methods	extrinsic	extrinsic	extrinsic	*ANAE*

Vowel intrinsic/ formant intrinsic (type 1)

- Psychoperceptual Transform Scales (Bark, ERB, mel etc)
- approximate the non-linear frequency response of the inner ear
- much more sensitive to changes in frequency at the lower end than higher in the spectrum.
- Bark units: one critical bandwidth is 100 Hz in the range between 150 and 250 Hz, but 350 Hz between 2150 and 2500 Hz
- Valid procedure to use a PTS and THEN a Normalization algorithm
- And not the other way around!

Vowel intrinsic/formant extrinsic (type 2)

- Bark Difference Metric
- Syrdal and Gopal (1986)

- Two slightly different versions of same idea:
- Hertz values are converted into Bark
- Z_3-Z_2 or Z_2-Z_1 modelling advancement
- Z_1-Z_0 modelling vowel height (NORM uses Z_3-Z_1)

Vowel extrinsic/formant intrinsic (type 2)

- Adank's most successful category for sociophonetic purposes (Adank 2003)

- Ranges (Gerstman)
- Z-scores (Lobanov)
- Individual log-means (Nearey CLIHi4)
- Centroids (Watt and Fabricius)

Vowel extrinsic/formant extrinsic (type 2)

- Nordström and Lindblom's (1975) vocal tract scaling transformation
- Nearey's shared log-mean model (Constant log Interval Hypothesis or CLIHs4/s2)
- Assuming a 'sliding template' approach
- Latter very common in N.Am studies

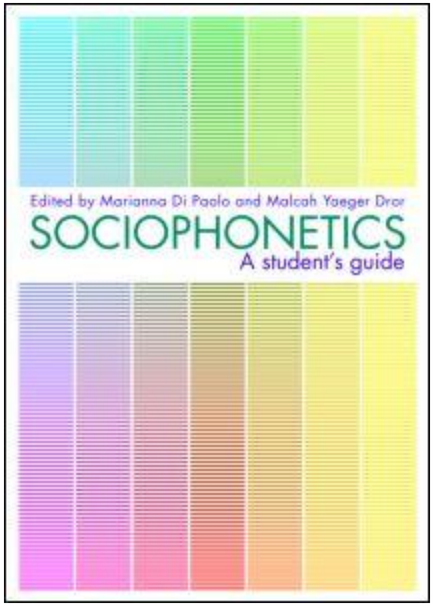
2 versions of Nearey

- Nearey 1 (NORM): CLIHi2 is formant-intrinsic
- Nearey 2 (NORM): CLIHs2 is formant-extrinsic
- Adank (2003) rates CLIHi4 as most successful (same typology as Lobanov)

More on Normalization and Plotting

- Watt, Fabricius and Kendall Chapter on these topics in

Sociophonetics: A Student's Guide



Edited by **Marianna Di Paolo, Malcah Yaeger Dror**

Part 2: What did that algorithm do?

- Fabricius, Watt and Johnson forthcoming, in LVC October 2009

A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics

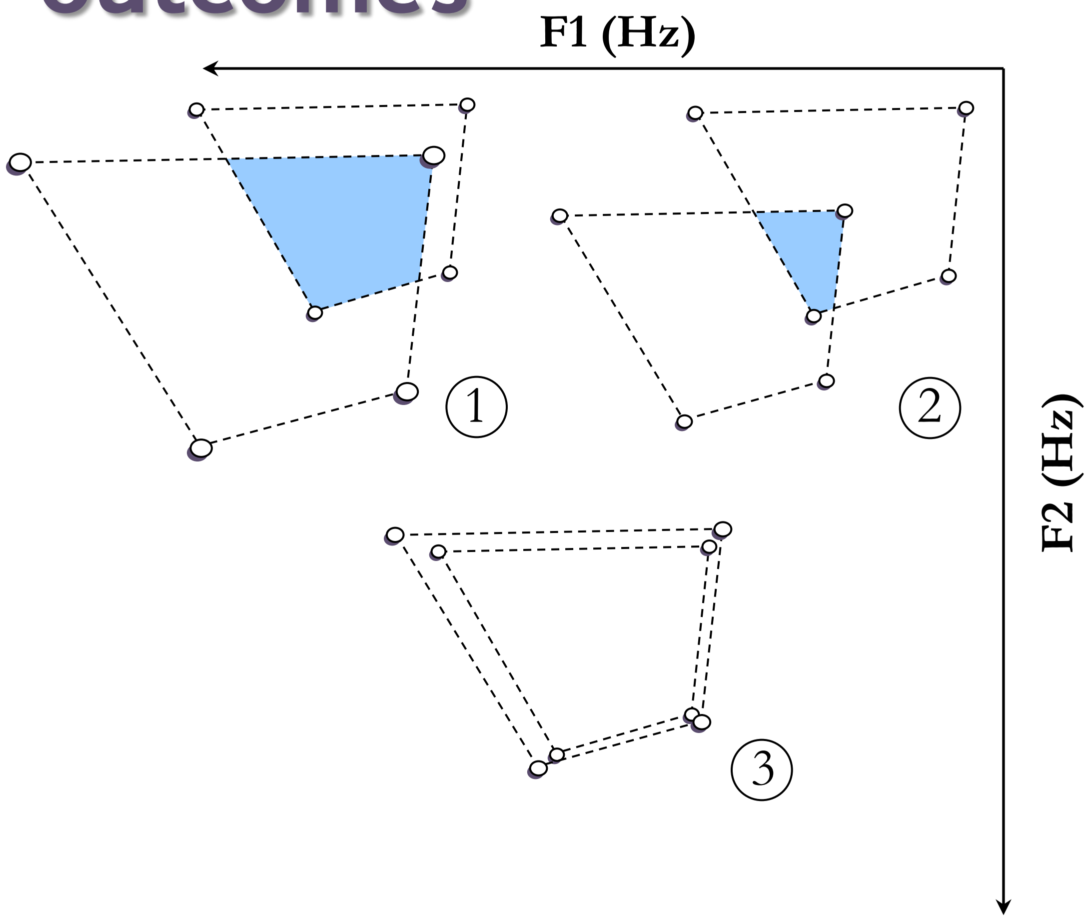
Motivation

- To test a (relative newcomer) normalization method based on vowel space centroids (Watt and Fabricius 2002)
- As it is typologically identical to Adank's 2003 most successful category
- **Vowel extrinsic and formant intrinsic**

Our research questions

- How do the *S*-procedures (*W&F*, and *mW&F*) perform compared with *Lobanov* and *Nearey1* on the following sociophonetically-relevant evaluative parameters
  - Reduction of variance in area ratios of vowel spaces, equalizing vowel space **areas (1)** as far as possible;
  - Improvement of **intersection (2)** of vowel polygons;
  - Reproduction of two-dimensional **vowel configurational relationships (3)** within the vowel space, compared with raw Hertz formant data

Schematic examples of normalization outcomes

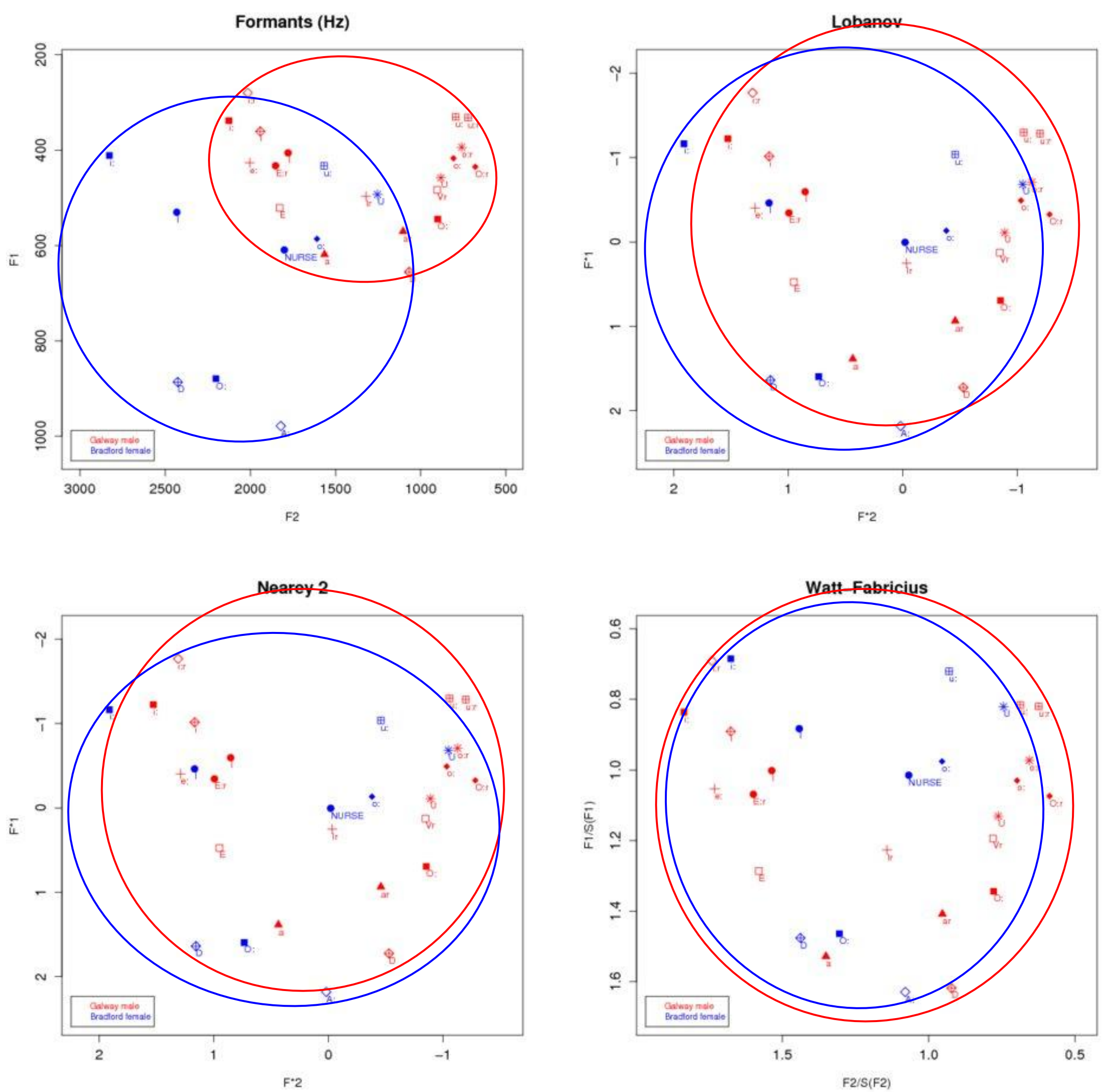


- 1 = disparity in area agreement and poor overlap
- 2 = good area agreement but poor overlap
- 3 = good fit on both counts

Normalisation

Galway male

Bradford female

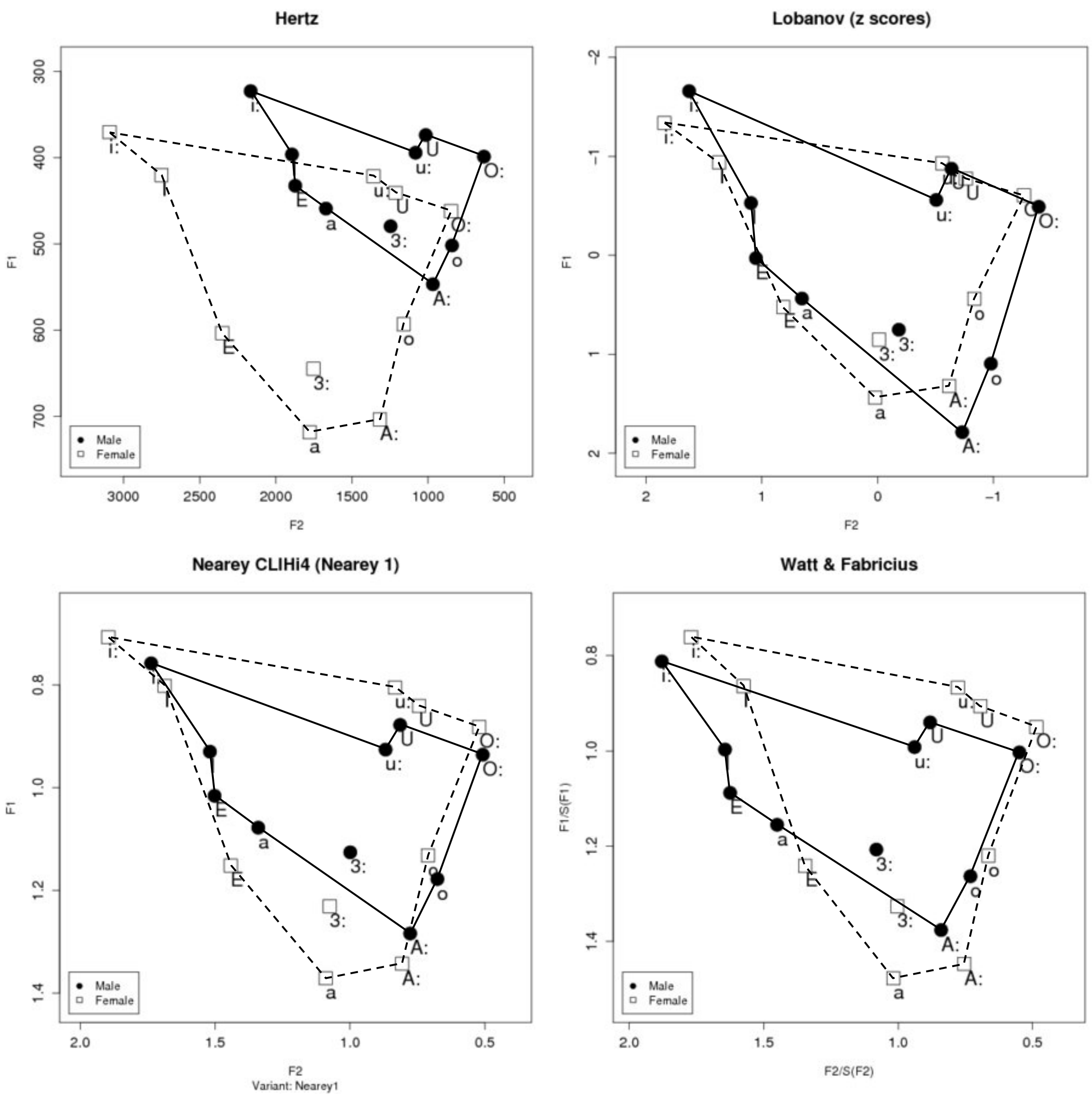


$F_{n[V]}^{N} = (F_{n[V]} - \bar{x}_n)/S_n$

$F^{*}_{n[V]} =$ anti-log$(\log(F_{n[V]}) - \bar{x}_{log})$

$S(F_n) = ([i]F_n + [a]F_n + [u']F_n)/3$

Two RP examples



Aims

- To test on larger data sets (RP: 20 speakers, Aberdeen: 6 speakers)
- Test statistically
- Establish procedures for comparisons on sociophonetically relevant **visual** parameters

- (cf Adank's 2003 study measures of sorting using linear discriminant analyses)

Our methods

- Area (1)
  - proportional reduction in variance
  - Pitman-Morgan's test of homogeneity of variance between correlated samples (Cohen, 1990).
- Intersection (2)
  - intersection of two vowel polygons divided by the union of the same polygons = intersection values compared statistically
- Vowel juxtapositions (3)
  - Angle calculations compared across methods

Our results: Test 1

Improvement over Hertz	*Nearey1*	*W&F*	*mW&F*	*Lobanov*
RP	0.070	0.350	0.389	0.923
Aberdeen	0.601	0.877	0.865	0.974

Table 2: Performance of normalization algorithms in removing variance between vowel space areas; higher values indicating a better result

Nearey 2: RP 0.071, Aberdeen 0.670

Lobanov ≥ W&F, mW&F > Nearey1 > Hertz

Our results: Test 2

Vowel space overlaps	*Hertz*	*Nearey1*	*W&F*	*mW&F*	*Lobanov*
RP average	.380	.444	.452	.500	.564
Aberdeen average	.444	.571	.598	.618	.688

Table 4: Average vowel space overlaps between any single speaker's vowel space and all other speakers' vowel spaces, RP and Aberdeen data sets.

Nearey 2: RP .445, Aberdeen .583

Lobanov > mW&F ≥ W&F, Nearey1 > Hertz

Our results: Test 3 (1)

Average angle values	*Hz*	*Nearey1*	*W&F*	*Lobanov*
TRAP-STRUT relative to horizontal, older speakers	2	5	5	4
TRAP-STRUT relative to horizontal, younger speakers	41	66	66	67
LOT-FOOT relative to vertical, older speakers	32	15	14	11
LOT-FOOT relative to vertical, younger speakers	81	65	66	65

Our results: Test 3 (2)

TRAP-STRUT Angle relative to horizontal	*Hz*	*Nearey1*	*W&F*	*Lobanov*
Mean differences between older and younger groups	38	61	61	64
Standard Deviation	12	25	26	30
LOT-FOOT Angle relative to vertical	*Hz*	*Nearey1*	*W&F*	*Lobanov*
Mean differences between older and younger groups	49	51	52	54
Standard Deviation	24	13	12	9

Table 8: Mean differences in angles, older speakers compared to younger speakers

Read more in:

Anne Fabricius, Dominic Watt and Daniel Ezra Johnson. 2009.

A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics.

Language Variation and Change. 21,3: 1-23.

Part 3: What do I see now?

- Based on publications Fabricius 2007a and 2007b, 2007c
- TRAP STRUT and LOT FOOT in the short vowel system of modern RP (UK) over time
- Introduced Angle juxtaposition calculations based on vowel plots
- *Advantages: mathematical precision and replicability*

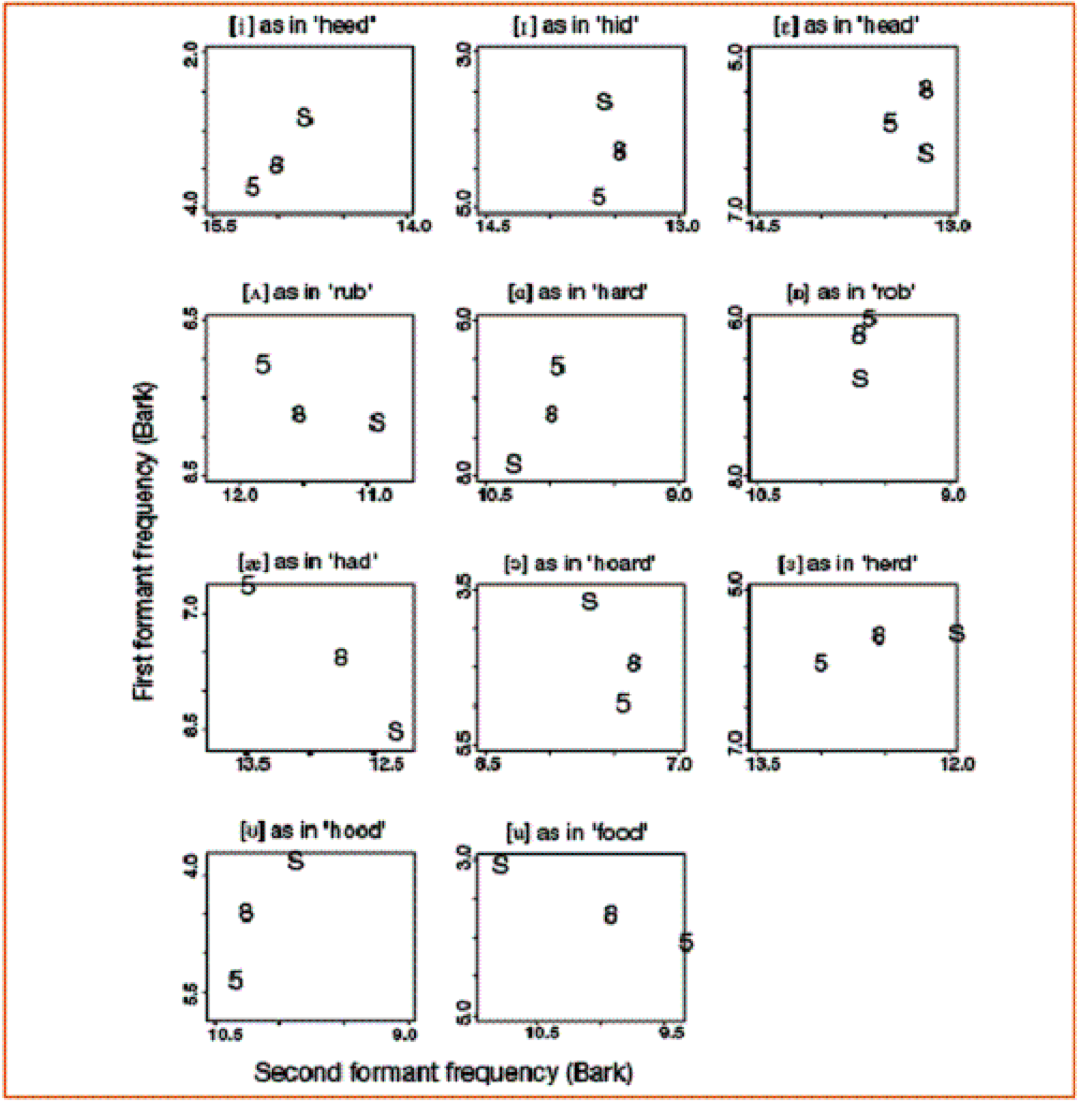


Figure 1 The three symbols '5', '8' and 'S' represent the average positions of different vowel types in the Christmas broadcasts of the 1950s and 1980s, and in standard southern British of the 1980s, respectively. The axes are the first two formant frequencies in Bark, a scale used to model the way listeners perceive vowels[12]. Positions towards the top of each square correspond to less mouth opening; the left corresponds to sounds made by constricting the vocal tract nearer the lips rather than further back[11].

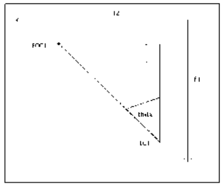
Acoustic phonetics/Sociophonetics

- Harrington et al as typical acoustic phonetic study
- Sociophonetics differs subtly

- Visual comparisons on more than one axis at a time

- Can we make these visual comparisons more stringent by using mathematical methods?

Methods



The juxtaposition of TRAP and STRUT relative to the horizontal

(1)   Tan Θ = ((F1 TRAP-F1 STRUT)/(F2 TRAP-F2 STRUT))

Juxtaposition of LOT and FOOT relative to the vertical:

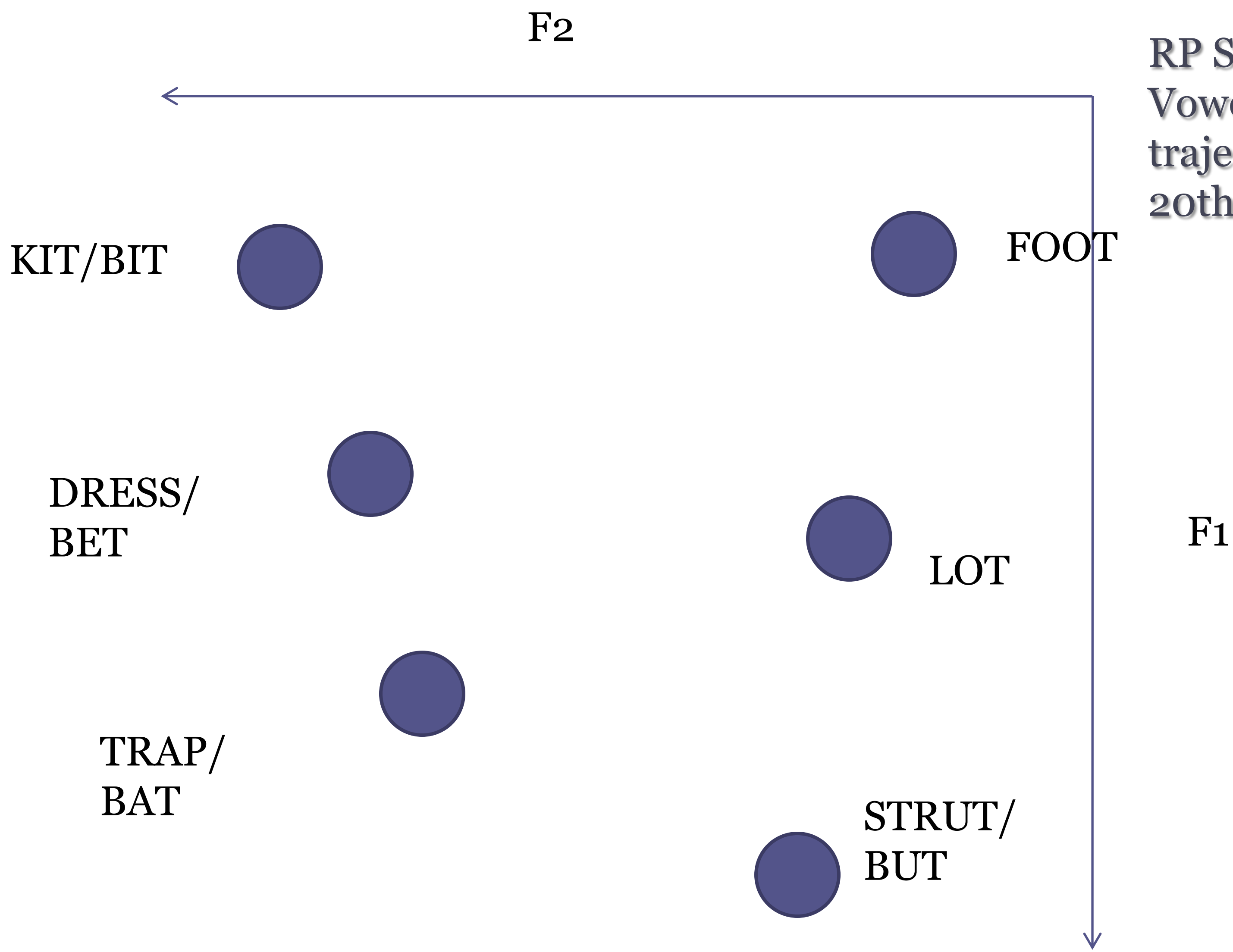
(2)   TAN Θ = ((F2 FOOT-F2 LOT)/(F1 LOT-F1 FOOT))

The tan Θ can derive the value of the angle in radians, for example using Microsoft Excel's ATAN (= Arctan) function.

Excel's DEGREE function then converts the angle from radians to degrees.

(3) DISTANCE (x,y) = $\sqrt{(F1\,x - F1\,y)^2 + (F2\,x - F2\,y)^2}$

RP Short Vowel trajectories in 20th century



Results

- Variation over time
- Configurations vary
- oldest and youngest generations show clearest pattern
- Middle groups vary;

Some individuals are leaders, others lag

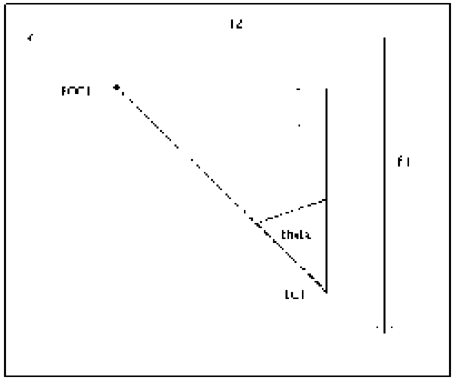
Born	TRAP/ STRUT	Euc.dist.	LOT / FOOT	Euc.dist.	Data Source
1909	-44	0,376	1	0,525	A, male h
1926	-18	0,314	27	0,158	D, 1950s
1926	2	0,249	13	0,340	D, 1960s
1926	-11	0,211	10	0,376	D, 1980s
1927	6	0,241	36	0,285	A, male c
1928-1936	-24	0,447	25	0,322	C, Sp.1-1
1928-1936	19	0,434	5	0,275	C, Sp.1-2
1928-1936	14	0,285	-2	0,226	C, Sp.1-3
1928-1936	9	0,496	17	0,585	C, Sp.1-4
1928-1936	18	0,254	22	0,334	C, Sp.1-5
Bef. 1945	10	0,352	5	0,508	B
1946-1951	8	0,230	6	0,553	C, Sp.2-1
1946-1951	47	0,341	6	0,314	C, Sp.2-2
1946-1951	-21	0,282	1	0,420	C, Sp.2-3
1946-1951	-27	0,444	27	0,212	C, Sp.2-4
1946-1951	67	0,615	35	0,306	C, Sp.2-5
1956	41	0,347	*78*	*0,068*	E
1961-1966	62	0,616	3	0,391	C, Sp.3-1
1961-1966	-26	0,264	-1	0,289	C, Sp.3-2
1961-1966	-2	0,405	13	0,426	C, Sp.3-3
1961-1966	40	0,331	*56*	*0,072*	C, Sp.3-4
1961-1966	27	0,399	61	0,214	C, Sp.3-5
1966	52	0,221	73	0,148	E
1973	*119*	*0,098*	57	0,244	E
1976-1981	61	0,396	69	0,202	C, Sp.4-1
1976-1981	84	0,637	64	0,126	C, Sp.4-2
1976-1981	56	0,624	68	0,104	C, Sp.4-3
1976-1981	79	0,823	58	0,162	C, Sp.4-4
1976-1981	58	0,353	62	0,127	C, Sp.4-5
1980	70	0,379	65	0,191	E

Conclusions

- Using angle calculations of this type sharpens the **replicability** of observed juxtapositional variations
- There is a place for

mathematical precision

in describing
chain shifts, splits, and
mergers over time……



Bibliography I

- Adank, Patti. 2003. *Vowel Normalization: A Perceptual-acoustic Study of Dutch Vowels*. PhD thesis, University of Nijmegen.
- Fabricius, Anne H. 2007a. Vowel Formants and Angle Measurements in Diachronic Sociophonetic Studies: FOOT-fronting in RP. *Proceedings of the 16th ICPhS, Saarbrücken*.
- Fabricius, Anne H. 2007b. Variation and Change in the TRAP and STRUT vowels of RP: a real time comparison of five acoustic data sets. *Journal of the International Phonetics Association* 37.3, 293-320.
- Fabricius, Anne H. 2007c. Using angle calculations to demonstrate vowel shifts: A diachronic investigation of the short vowel system in 20th-century RP (UK). *Acta Linguistica Hafniensis*.
- Gerstman, Louis. 1968. Classification of self-normalized vowels. *IEEE Transactions of Audio Electroacoustic*s AU-16: 78-80.
- Lobanov, Boris M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49(2B):606–608.
- Nearey, Terrance. 1977/8. *Phonetic Feature Systems for Vowels*. Indiana University Linguistics Club. [Online: <http://www.ualberta.ca/~tnearey/Nearey1978_compressed.pdf>]
- Nearey, Terrance and Peter Assmann. 2007. Probabilistic 'sliding template' models for indirect vowel normalization. In Maria-Josep Solé, Patrice Beddor and Manjari Ohala (eds.) *Experimental Approaches to Phonology*. Oxford: Oxford University Press, pp. 246-269.

Bibliography II

- Nordström, P.-E. and Björn Lindblom. 1975. A Normalization Procedure for Vowel Formant Data. *Proceedings of the 8th International Congress ofPhonetic Sciences*, Leeds, p. 212.
- Pisoni, David. 1997. Some thoughts on "normalization" in speech perception. In Keith Johnson and John Mullennix (eds.) *Talker Variability in Speech Processing*. San Diego: Academic Press, pp. 9-32.
- Stevens, Stanley and John Volkman. 1940. The relation of pitch to frequency: a revised scale. *American Journal of Psychology* 53: 329-353.
- Thomas, Erik R. and Tyler Kendall. 2007. *NORM: The Vowel Normalization and Plotting Suite*. [Online: <http://ncslaap.lib.ncsu.edu/tools/norm/>]
- Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88: 97-100.
- Traunmüller, Hartmut. 1997. *Auditory Scales of Frequency Representation*. [Online: <http://www.ling.su.se/staff/hartmut/bark.htm>]
- Watt, Dominic and Anne Fabricius. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1~F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9: 159-73.

Acknowledgements

- My co-authors:
  - Dominic Watt
  - Daniel E. Johnson
  - Tyler Kendall
- Malcah Yæger-Dror and Marianna di Paolo
- Department of Culture and Identity, Roskilde University

Thank you!

This presentation is available at the following website:

http://akira.ruc.dk/~fabri/nwav38.pdf