Analysis of vowel formants vs. spectral-shape features in real-time studies of language variation and change

Tamara Rathcke[1], Jane Stuart-Smith[1], Brian José[1], Claire Timmins[2], Bernard Torsney[1]

[1]University of Glasgow, [2]University of Strathclyde, United Kingdom

GULP

The Leverhulme Trust

Research motivation

Sociophonetic real-time studies look into the development of sound systems by comparing variation sampled at different time points (e.g. Sankoff 2006; Meyerhof 2006).

They allow for great insights into individual patterns of variation and actual rates of change (Sankoff & Blondeau 2007; Gregersen, Maegaard & Pharao 2009; Labov 1994).

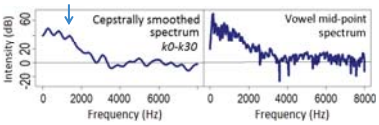
Methodologically "ideal" cases (exactly replicated recording situation and equipment, high quality recordings) are unusual. However, acoustic investigations of diverse real-time recordings may be compromised by several factors:

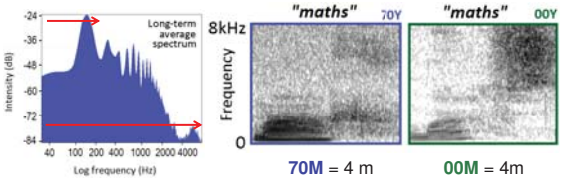
- room acoustics, ambient noise, placement and type of the microphone all have an impact on acoustic signals (Brixen 1996; Decker & Nycz 2010; Hansen & Pharao 2006, subm.; Plichta 2004)
- LPC-algorithms used for formant measurements are very sensitive to even slight changes in spectral properties such as signal-to-noise ratios, resonance frequencies, spectral balance. Especially F1 is affected. The amount of deviation differs for each vowel category.

How can we ascertain that acoustic differences we observe in real-time corpora (e.g. /u/-fronting) are genuine, not due to technical artefacts?

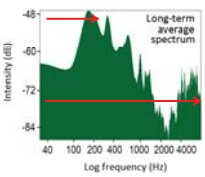
A real-time corpus of Glaswegian vernacular

1970s-recordings (Macaulay 1977):
- quiet rooms
- *Uher M822* lavalier microphone
- (not always) placed on the speaker
- downsampled to 16kHz

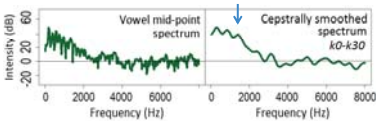




70M = 4 m
70 Y = 4 m
00M = 4m
00Y = 4m



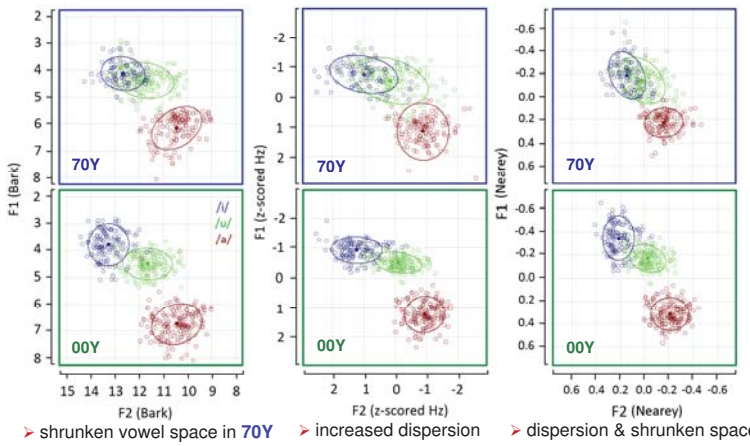
2000s-recordings (Stuart-Smith 2006):
- quiet rooms/noisy pubs
- *Sony ECM CS10* lavalier microphone
- placed on the speaker's chest
- downsampled to 16kHz



Large spectral discrepancies diminish after cepstral smoothing. Are spectral-shape features a more reliable source of information?

Formant analyses

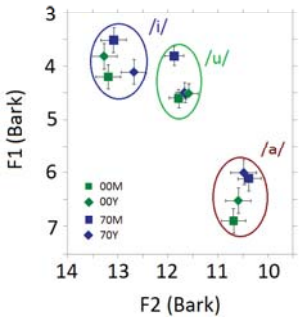
F1/F2-vowel space as created by Bark- vs. Lobanov- vs. Nearey-transforms



- shrunken vowel space in **70Y**
- increased dispersion
- dispersion & shrunken space

Normalisation by estimation in LMEM

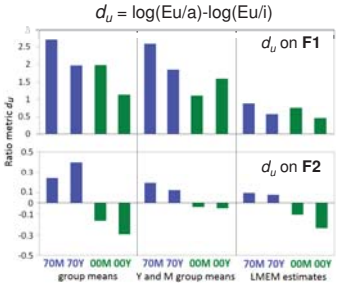
lmer(Fn ~fixed factors + (1|speaker+(1|word)+(1|decade)+random slopes)



- vowel spaces more comparable across the two decades of recording, esp. F2
- F1-spread remains & F1/F2-overlap of /i/ and /u/ in **70Y** (and **70M**) disappears

Normalisation by relative positioning of /u/ between /i/ and /a/ (after Harrington et al. 2008)

d_u = log(Eu/a)-log(Eu/i)

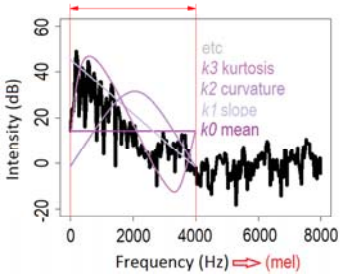


- relative positioning of /u/ is changing towards /a/
- "rate of change" depends on reference centroids, F2 more reliable than F1

Relative metrics (and linear mixed-effects modelling) deal better with technical issues than commonly used vowel transformations.

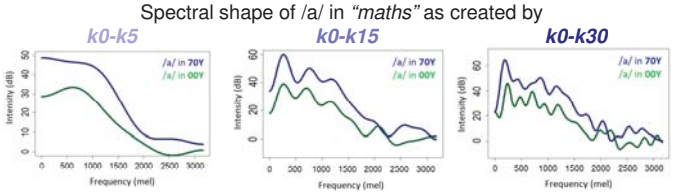
Spectral-shape analyses

Procedure:
- frequency range 0-4000 Hz
- Mel-scaled
- vowel mid-points
- decomposition in cepstral coefficients (discrete cosine transformation, DCT)

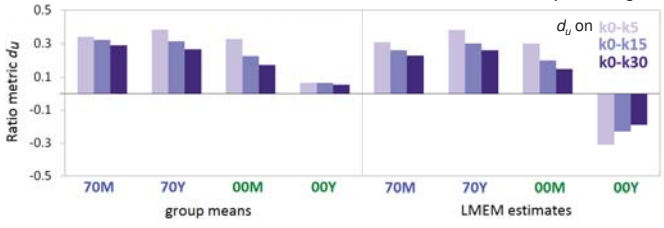


- at least 6 DCT-coefficients are needed to reflect information similar to F1/F2, *k0-k5*

Spectral shape of /a/ in *"maths"* as created by

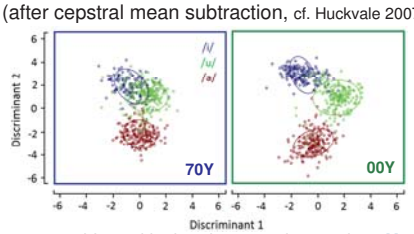


- the more coefficients included, the more information about vowel spectrum given

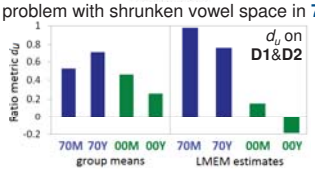


- d_u-output is comparable to F1/F2-patterns (potentially includes F3-information)
- number of coefficients does not have a very strong impact on the d_u-output

Linear discriminant analysis on *k0-k15* (after cepstral mean subtraction, cf. Huckvale 2007)



- problem with shrunken vowel space in **70Y**



Distance metric d_u is more robust when applied to DCT-smoothed spectra than to formants but technical problems remain.