On the impact of noise on vowel formants

Tamara Rathcke* and Jane Stuart-Smith+

*University of Kent;  +University of Glasgow

Special Session: ‘Of vowels and “systems”’

Methods XV, University of Groningen, 11 August 2014
Researching sound change

• *Structural changes in the sound system of a language over time*
  – as opposed to synchronic variability, e.g. style, region, social, individual etc.

• Main goals
  – What are the patterns of change?
  – How does sound change arise?
Observing sound change

• ‘apparent-time’ (e.g. Labov 1972; Labov 1994)
  – Recordings of speech made at one point in time, from several age groups
  – Relative stability of variation across lifespan (cf Sankoff and Blondeau 2007)
  – The issue of age-related variation at specific life stages (‘age-grading’) e.g. Sankoff (2006)
Observing sound change

• ‘real-time’ (e.g. Labov 1994)
  – Recordings of speech made at several points in time, from one or more groups
  – Comparability of data in real-time recordings (e.g. Tillery and Bailey 2003; Gregersen et al this conference).
  – ‘ideal’ corpora, i.e. with exactly the same recording situation, are rare (?impossible)
Comparability in the acoustic dimension

- Room acoustics, background noise, make and placement of microphones, inter alia, have an effect on recording quality (Brixen 1996; Plichta 2004; Hansen & Pharao 2006)

- LPC algorithms are very sensitive to variability in recording quality, especially Signal to Noise Ratio (SNR) (Plichta 2004)

- F1 is particularly vulnerable (Moye 1979, Künzel 2001, Byrne & Foulkes 2004; Hansen & Pharao under revision)
e.g. microphones and formant frequencies

<table>
<thead>
<tr>
<th></th>
<th>MKH</th>
<th>DPA</th>
<th>VT700 DPA</th>
<th>VT700 VT700</th>
</tr>
</thead>
<tbody>
<tr>
<td>rel. to</td>
<td>B&amp;K</td>
<td>B&amp;K</td>
<td>B&amp;K</td>
<td>MKH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MKH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DPA</td>
</tr>
<tr>
<td>Semitones</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>1.319</td>
<td>2.457</td>
<td>2.191</td>
<td>1.138</td>
</tr>
<tr>
<td>F2</td>
<td>-0.223</td>
<td>-0.189</td>
<td>-0.017</td>
<td>0.035</td>
</tr>
<tr>
<td>F3</td>
<td>0.868</td>
<td>-1.193</td>
<td>-0.997</td>
<td>-0.325</td>
</tr>
</tbody>
</table>

Brüel & Kjaer 4179

Sennheiser MKH40

DPA 4066

Voice Technologies VT700

Hansen & Pharao, under revision
What kind of impact do technical differences in recordings have on formant analyses for real-time investigation?
Fine phonetic variation and sound change: A real-time study of Glaswegian

http://soundsofthecity.arts.gla.ac.uk/

Oct 2011-December 2014
(target) corpus for a real-time study of Glaswegian Vernacular

<table>
<thead>
<tr>
<th>Decade of Recording</th>
<th>Old 67-90 (Decade of Birth)</th>
<th>Middle-aged 40-55 (Decade of Birth)</th>
<th>Young 10-15 (Decade of Birth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970s</td>
<td>6 f, 6 m (1890s)</td>
<td>6 f, 6 m (1920s)</td>
<td>6 f, 6 m (1960s)</td>
</tr>
<tr>
<td>1980s</td>
<td>6 f, 6 m (1900s)</td>
<td>6 f, 6 m (1930s)</td>
<td>6 f, 6 m (1970s)</td>
</tr>
<tr>
<td>1990s</td>
<td>6 f, 6 m (1910s)</td>
<td>6 f, 6 m (1940s)</td>
<td>6 f, 6 m (1980s)</td>
</tr>
<tr>
<td>2000s</td>
<td>6 f, 6 m (1920s)</td>
<td>6 f, 6 m (1950s)</td>
<td>6 f, 6 m (1990s)</td>
</tr>
</tbody>
</table>

Private, force-aligned, electronic corpus of spontaneous speech, available for academic researchers, covering 100 years of Glaswegian in apparent- and real-time, stored in LABB-CAT (Fromont and Hay 2012).
Sample for this paper – 24 male speakers

<table>
<thead>
<tr>
<th>Decade of Recording</th>
<th>Old 67-90 (Decade of Birth)</th>
<th>Middle-aged 40-55 (Decade of Birth)</th>
<th>Young 10-15 (Decade of Birth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970s</td>
<td>4m (oral history interview)</td>
<td>4m (sociolinguistic interview)</td>
<td>4m (sociolinguistic interview)</td>
</tr>
<tr>
<td>1980s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000s</td>
<td>4m (oral history interview)</td>
<td>4m (conversation)</td>
<td>4m (conversation)</td>
</tr>
</tbody>
</table>

Sources (with thanks): Ronald Macaulay; William Labov; Paul Thompson; Glasgow Media Project; Glasgow People’s Palace
good...

- 70-O, m3: "good thing"
- 70-M, m4: "good old days"
- 70-Y, m2: "good place"
- 00-O, m3: "good people"
- 00-M, m1: "good for me"
- 00-Y, m3: "good fuels"
SNR = Signal-to-Noise Ratio

\[ SNR = \frac{P_{signal}}{P_{noise}} = 20 \log_{10} \left( \frac{A_{signal}}{A_{noise}} \right) \]
Differences in recording quality in the sample
Signal to Noise Ratio (SNR)

$p = 0.0056 \approx 0.05/9$

better
(less noise)

worse
(more noise)

“real-time” comparison
Differences in recording quality in the sample
Signal to Noise Ratio (SNR)

$p=0.0056 \ (\approx 0.05/9)$

better
(less noise)

worse
(more noise)

“real-time” comparison
“apparent-time” comparison
SB = Spectral Balance
SB = Spectral Balance
Differences in recording quality in the sample Spectral Balance (SB)

$p = 0.0056 \approx 0.05/9$

![Graph showing differences in recording quality.]

**better** (more balanced)

**worse** (less balanced)

“real-time” comparison
Differences in recording quality in the sample
Spectral Balance (SB)

$p = 0.0056 \approx 0.05/9$

"real-time" comparison
"apparent-time" comparison

better (more balanced)

worse (less balanced)
Research questions

1. Does recording quality have an impact on formant analysis?
2. Are all vowels affected in the same way?
3. Is it possible to reduce, or even completely remove, the complicating effects of SNR/SB?
Acoustic analysis of vowels

• accented /i a ʉ/ (no words with postvocalic /r/)
• N = 3610
• labelled in EMU

• Formant measures taken in Praat
  – downsampling to 10 kHz, low pass filter 5 kHz, LPC-order 10; pre-emphasis 50 Hz
• Manual correction of unreliable data points (F1: +/- 60 Hz; F2: +/- 450 Hz)
• Removal of outliers (defined in Bark F1 & F2, for each vowel)
  ➢ N = 3032 (16% removed)
Bark-transformed ‘raw’ data
Bark-transformed ‘raw’ data

Dispersion ($A_e$)
Bark-transformed ‘raw’ data

Dispersion ($A_e$)  Size of the vowel space between /i/, /a/ and /ʉ/ ($A_t$)

- **70O**
  - $n_i = 203$
  - $n_u = 189$
  - $n_a = 243$

- **70M**
  - $n_i = 124$
  - $n_u = 155$
  - $n_a = 157$

- **70Y**
  - $n_i = 95$
  - $n_u = 159$
  - $n_a = 132$

- **00O**
  - $n_i = 283$
  - $n_u = 189$
  - $n_a = 404$

- **00M**
  - $n_i = 65$
  - $n_u = 84$
  - $n_a = 77$

- **00Y**
  - $n_i = 141$
  - $n_u = 159$
  - $n_a = 171$
Recording quality and Bark-transformed vowel space

**Signal to Noise Ratio:**

– no correlation between SNR and the size of the vowel space between /i/-/a/-/ʉ/ ($A_t$)
– negative correlation for dispersion ($A_e$) of /ʉ/:  
  • the worse the SNR, the more dispersed the distribution ($R = -0.38$, $p=0.064$)
Recording quality and Bark-transformed vowel space

Spectral Balance:

— positive correlation between SB and the size of vowel space between /i/-/u/-/a/ ($A_t$):
  • the less balanced the spectrum, the smaller the space ($R = 0.46, p=0.025$)

— no correlation between SB and dispersion ($A_e$)
Influence of recording quality on formants

LME-analysis

- **Signal to Noise Ratio**
  - F1 of all vowels is 0.42 Bark higher in poorer SNR-environments (SNR: $\chi^2_{(1)} = 10.1$, $p=0.0015$)
  - no effect of SNR on F2
Influence of recording quality on formants

LME-analysis

• **Spectral Balance**
  
  – F1 of high vowels is lower in recordings with poorer spectral balance (vowel x SB: $\chi^2_{(2)} = 15.1, p < 0.001$)
  
  – F2 of high vowels is lower in recordings with poorer spectral balance (vowel x SB $\chi^2_{(2)} = 40.1, p < 0.0001$)
Is it possible to reduce the impact of recording quality on the vowel spectrum?
Suggestions

- Normalize the raw formant measures (e.g. Lobanov, Nearey, Watt-Fabricius)
- Use a distance metric (e.g. $d_u$, after Harrington, Kleber and Reubold 2008)
- ....
Formant normalization...

Nearey and Watt-Fabricius methods give the same results as raw Bark-transformed data:

- dispersion/size of vowel space:
  - The worse the SNR, the more dispersed the distribution of /ʉ/
  - The worse the SB, the smaller the space between the vowels

- same effects in the LME analyses:
  - F1 of all vowels is higher in poorer SNR environments
  - F1 and F2 of high vowels is lower with recordings with poorer SB
... an exception

Lobanov:

- no significant correlations between SNR/SB and $A_t/A_e$
- effect of SNR on F1 is removed
- effect of SB on F2 is removed
... an exception

Lobanov:

– no significant correlations between SNR/SB and $A_t/A_e$
– effect of SNR on F1 is removed
– but effect of SB on F1 remains for /i ɨ/
– effect of SB on F2 is removed
Distance metric: $d_u$

$$d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2)$$

after Harrington, Kleber and Reubold (2008:2828)
Distance metric: $d_u$

$$d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2)$$

after Harrington, Kleber and Reubold (2008:2828)
Distance metric: $d_u$

$$d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2)$$

after Harrington, Kleber and Reubold (2008:2828)
Distance metric: $d_u$

$$d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2)$$

after Harrington, Kleber and Reubold (2008:2828)
Distance metric: $d_u$

\[ d_u = \log(E_1/E_2) = \log(E_1) - \log(E_2) \]

No correlation between SNR/SB and $d_u$

after Harrington, Kleber and Reubold (2008:2828)
Return to the research questions

1. *Does recording quality have an impact on formant analysis?*
   - **YES**, both F1 and F2 are affected (Moye 1979, Künzel 2001, Byrne & Foulkes 2004; Hansen & Pharao under revision)

2. *Are all vowels equally affected?*
   - **NO**, in these data /i ʉ/ are more affected than /a/

3. *Is it possible to reduce, or even completely remove, the negative effects of SNR and SB?*
   - **YES**, it is possible to reduce it (Lobanov normalization; distance metric)
Signal to Noise Ratio and Spectral Balance are only two of several technical issues with recording quality...
e.g. reduced spectral information
e.g. reduced spectral information
e.g. reduced spectral information

recording made in 00s

recording made in 70s
... redigitization of reel-to-reel tapes with corrected position of tape-head

recording made in 70s before

recording made in 70s after

Many thanks to Gert Hansen for this insight
... and sound change?

- Raw Hz or Bark-transformed data
  - higher SNR predicts apparent vowel lowering
  - poorer SB predicts apparent vowel raising and retracting for high vowels
- With Lobanov normalization
  - poorer SB predicts apparent vowel raising for high vowels
- With distance metric
  - No predictions relating to recording quality
Group: $\chi^2_{(5)} = 12.6$, $p < 0.027$

/ʊ/ appears to be lowering over time (irrespective of SNR and SB)
Thank you!
and many thanks to
Brian José, Christoph Draxler, Anne Fabricius, Jonathan Harrington, Gert Hansen, Mark Huckvale, Tyler Kendall, Cordula Klein, Cerwyss O’Hare, Nicolai Pharao, Ulrich Reubold, Claire Timmins, Jim Scobbie, Michel Scheffers, Dominic Watt

The Leverhulme Trust