Modelling tongue position in German vowels

Philip Hoole
Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 Munich
Germany

“A hybrid PARAFAC and Principal Component model of tongue configuration in vowel production”
Two-Mode Principal Component Analysis

Two-dimensional data matrix:
  \( m \) observations (e.g. 15 vowels) on \( n \) variables (e.g. 8 articulators)

Each factor extracted represents a weighted sum of the 8 articulators
Each vowel has a score with respect to each factor

Problem: Rotational indeterminacy of factor axes
Three-Mode Analysis (PARAFAC)  
(e.g Harshman et al., 1977)

Systematic exploitation of a third dimension to solve the indeterminacy problem  
In this work the speakers represent this third dimension  
Analogy: Simultaneous equation  
  \[ x + y = 20 \]  
  \[ 2x + 3y = 55 \]

Model prediction for speaker \( k \):  
\[ Y_k = A S_k V^T \]
where \( V \), \( A \) and \( S \) are 3 loading matrices (for vowels, articulators and speakers, respectively), and where \( S_k \) is a matrix with the \( k \)th row of \( S \) on the main diagonal and zero elsewhere

Hence very strong assumptions on possible speaker-specific behaviour

If assumptions are met  
Very parsimonious representation  
Close relationship of factors to the underlying behavioural dimensions
Material

15 German vowels (monophthongs)
3 consonant contexts (pVp, tVt, kVk)
7 speakers
2 speech rates (separate recording sessions)
8 articulatory coordinates (x/y for 4 fleshpoints on tongue)

Preprocessing

Data averaged over 5 repetitions of each token
Data converted to deviations from each subject’s mean articulatory position
A bumpy road

A reliable 3-factor model could not be extracted

2-factor models:

<table>
<thead>
<tr>
<th>Contexts</th>
<th>RMS error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-context only</td>
<td>1.2</td>
</tr>
<tr>
<td>t-context only</td>
<td>!model unreliable!</td>
</tr>
<tr>
<td>k-context only</td>
<td>1.1</td>
</tr>
<tr>
<td>p and k contexts</td>
<td>1.5</td>
</tr>
<tr>
<td>p, t and k contexts</td>
<td>1.9</td>
</tr>
</tbody>
</table>

All the reliable 2-factor models were very similar to each other.
Tongue configuration: Factor 1

Y (mm)

hard palate

X (mm)
Tongue configuration: Factor 2

![Graph showing the configuration of the hard palate with different factors indicated by markers at specific X and Y coordinates.](image-url)
Physiological dimensions of vowel articulation
(S. Maeda and K. Honda, 1994)
PARAFAC subject loadings

Factor 1

Factor 2

Normal rate

Fast rate
Extending the model

Can the failure of the 3-factor model (and of the 2-factor t-context model) be explained?

Procedure:
Subject-specific principal-component analysis of the PARAFAC model error

Result:
The first principal component shows in all subjects an alternation between tongue-blade and tongue-dorsum raising
It explains about 50% of the variance
Boring brute-force approach

Simply retain for each subject the 45 vowel weights (15 vowels * 3 contexts) and the 8 articulator weights

RMS error 1.1mm

“Spirit of PARAFAC” approach

Use subject-specific articulator weights (as above)  
but
retain only 1 set of vowel weights (averaged over subjects)

RMS error 1.2mm
Combined PARAFAC and error-analysis model
Final Model

2 PARAFAC factors
1 factor derived from subject-specific principal component analysis of the PARAFAC model error
using
   vowel scores averaged over speakers
but
   speaker-specific articulator weights

Conclusions

1. The basic PARAFAC approach gives a succinct and physiologically plausible account of vowel articulation
2. Consonantal articulation requires a more complex subject-specific mapping between underlying articulatory component and observable fleshpoint coordinates