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*: $\mathbf{Y}_k=\mathbf{AS}_k\mathbf{V}^T$

where **V**, **A** and **S** are 3 loading matrices (for vowels, articulators and speakers, respectively), and where $\mathbf{S}_k$ 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

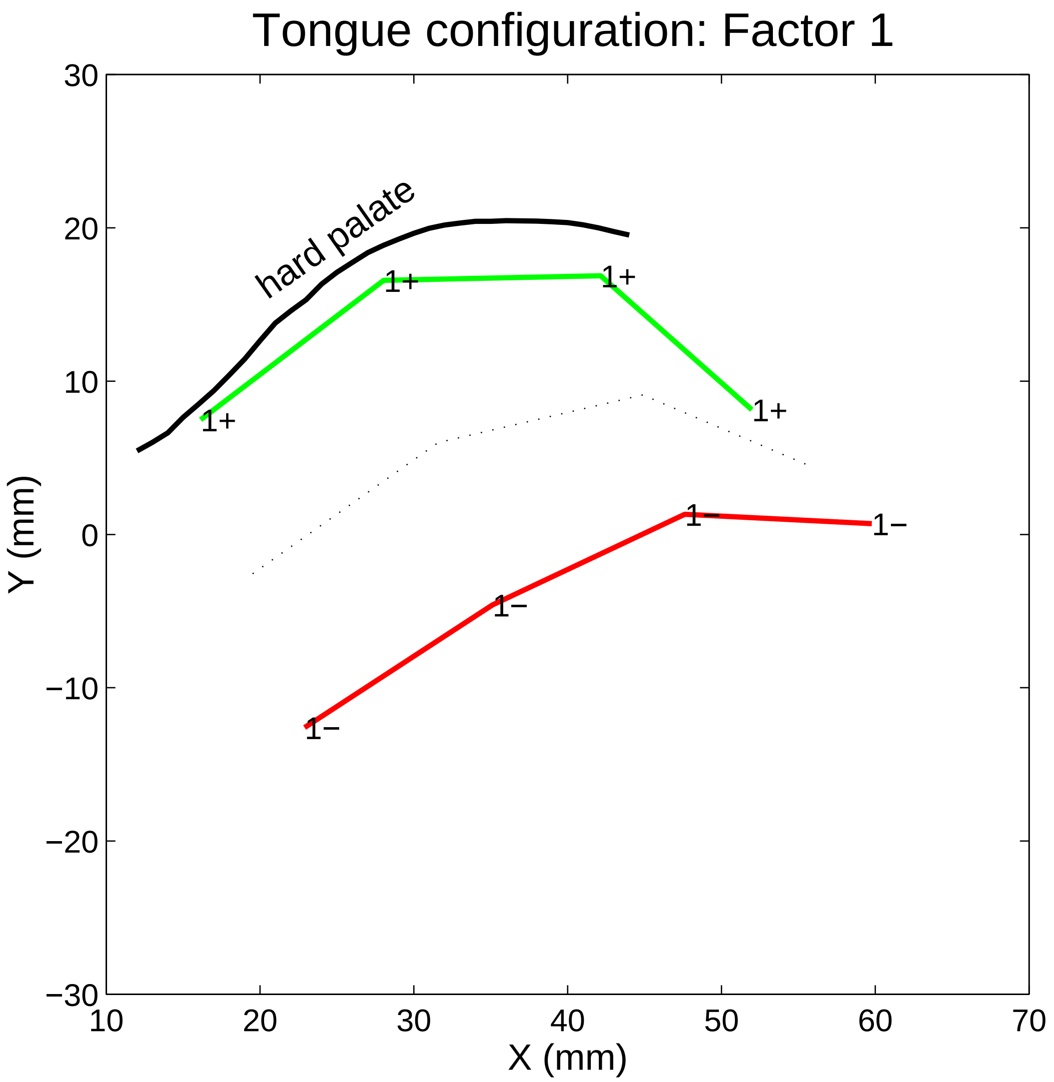
A bumpy road

A reliable 3-factor model could not be extracted

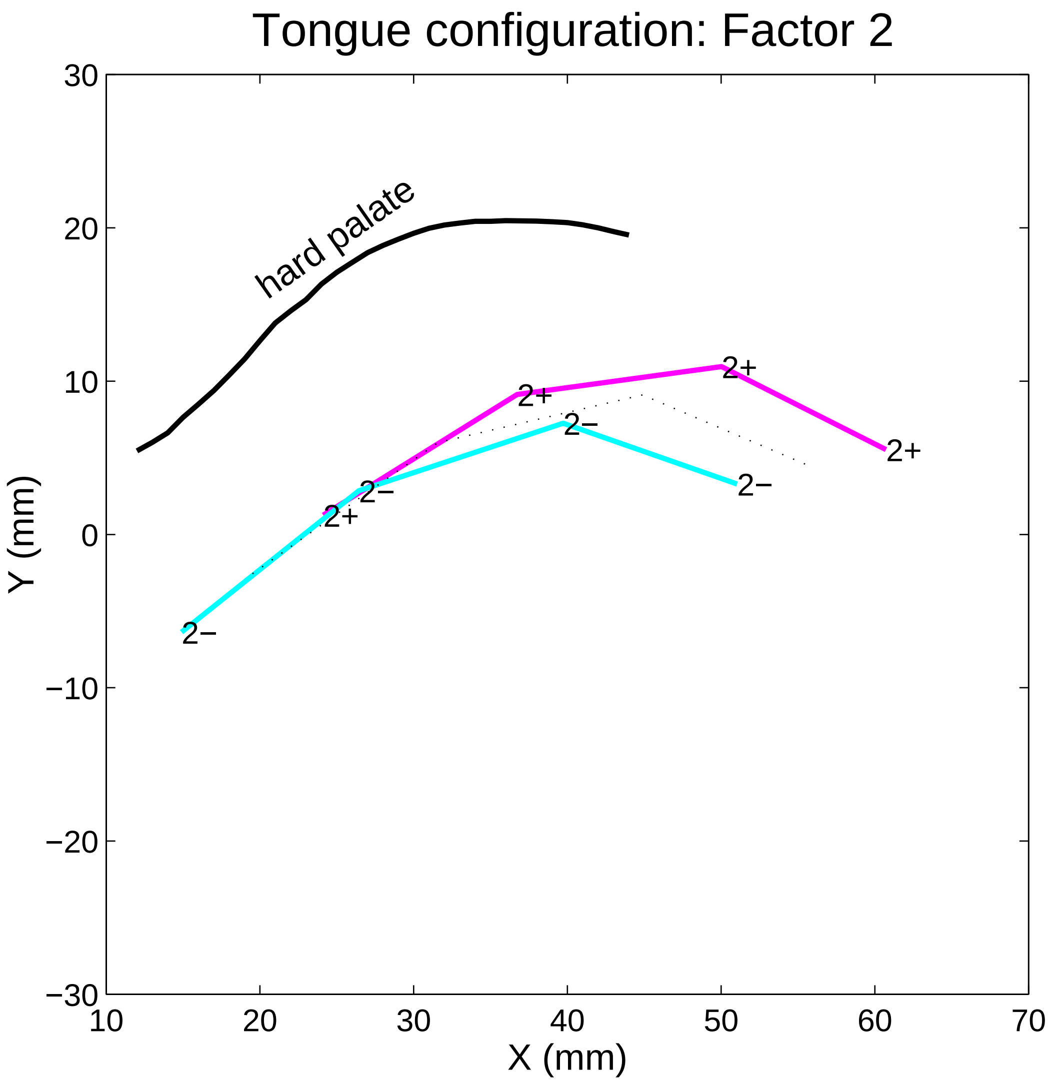
2-factor models:

	RMS error (mm)
p-context only	1.2
t-context only	!model unreliable!
k-context only	1.1
p and k contexts	1.5
p, t and k contexts	1.9

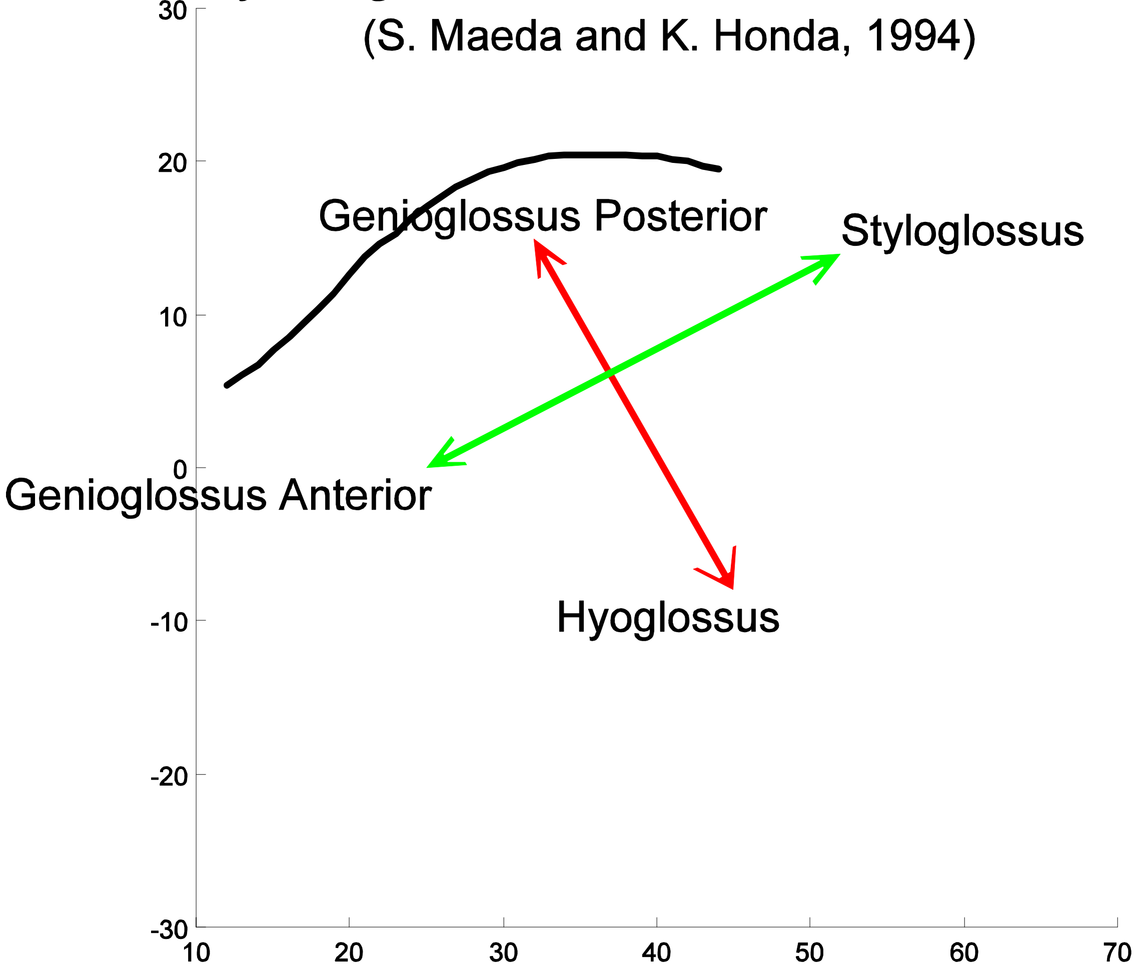
All the reliable 2-factor models were very similar to each other

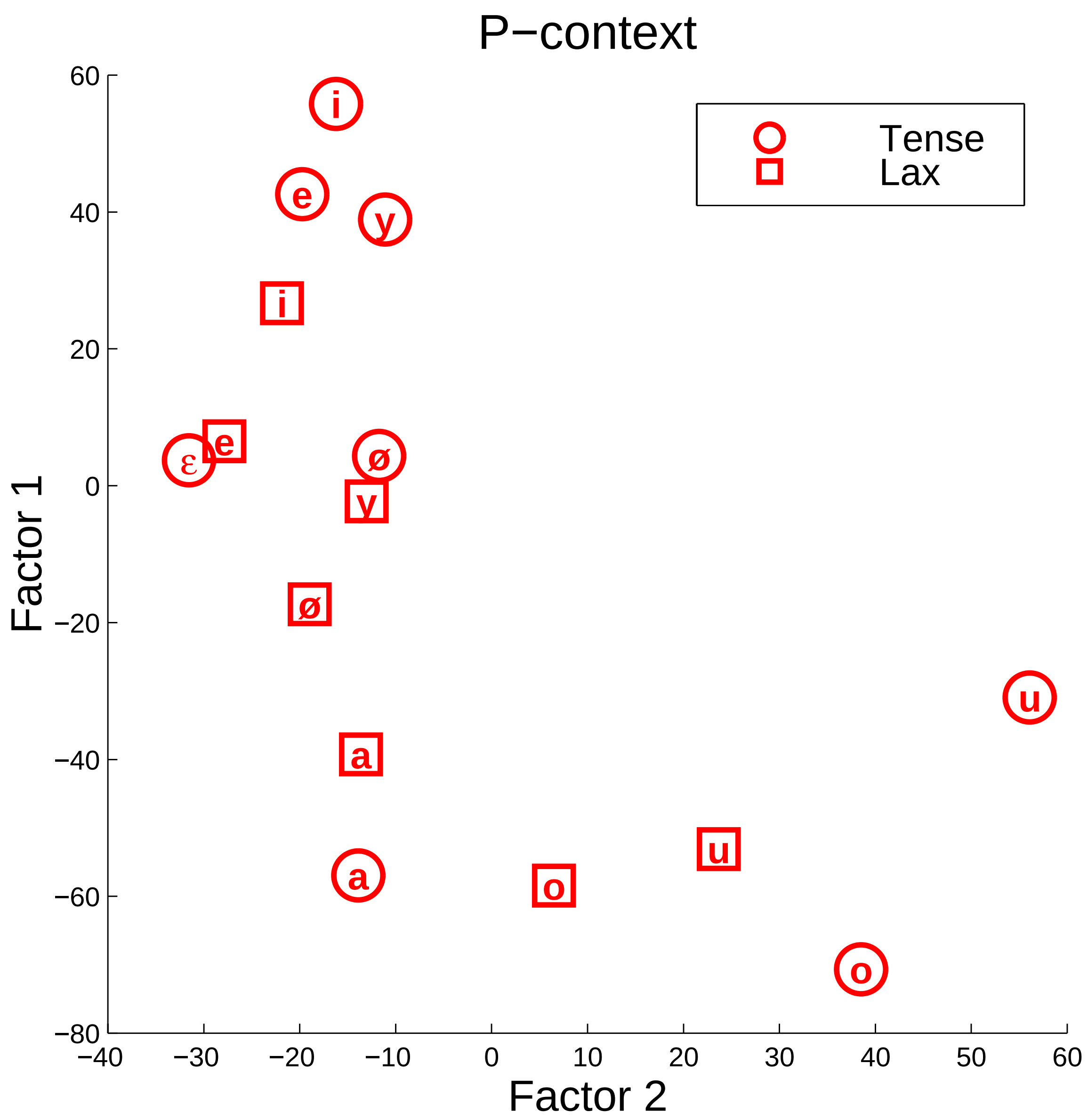


Tongue configuration: Factor 1
hard palate
1+
1+
1+
1+
1−
1−
1−
1−
Y (mm)
X (mm)
30
20
10
0
−10
−20
−30
10
20
30
40
50
60
70

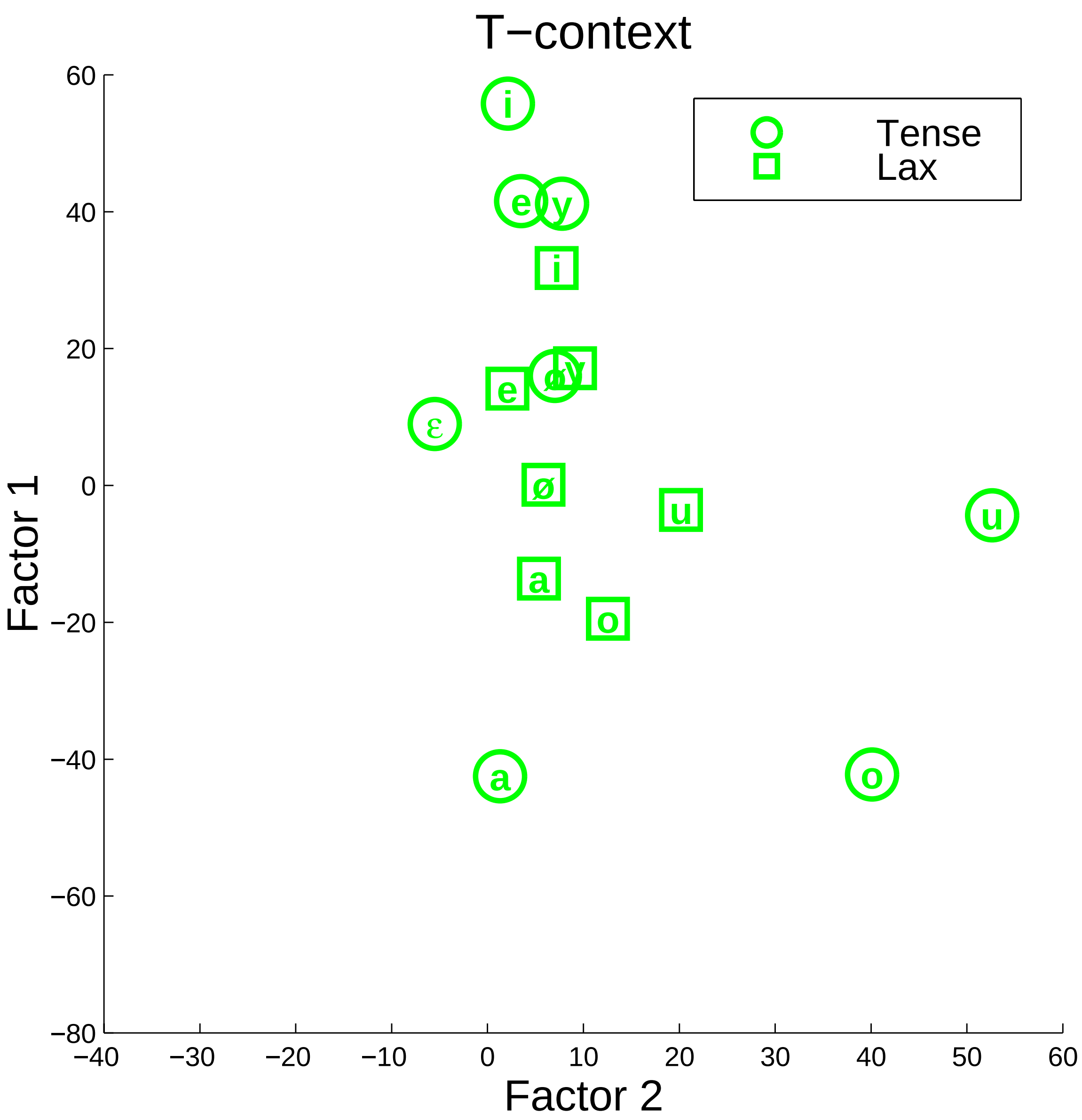

Tongue configuration: Factor 2
hard palate
2+
2+
2+
2+
2−
2−
2−
2−
Y (mm)
X (mm)
30
20
10
0
−10
−20
−30
10
20
30
40
50
60
70

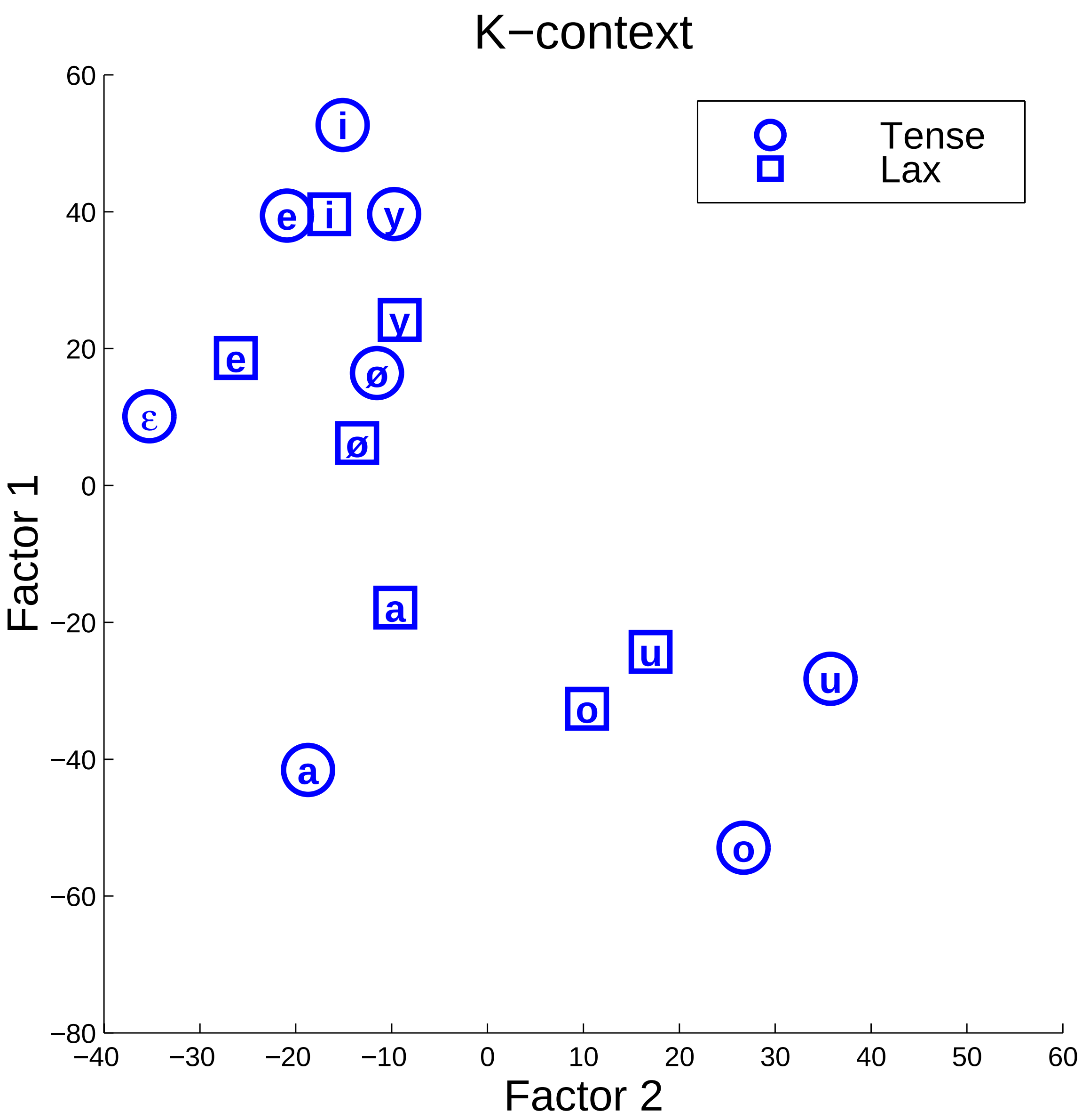
Physiological dimensions of vowel articulation

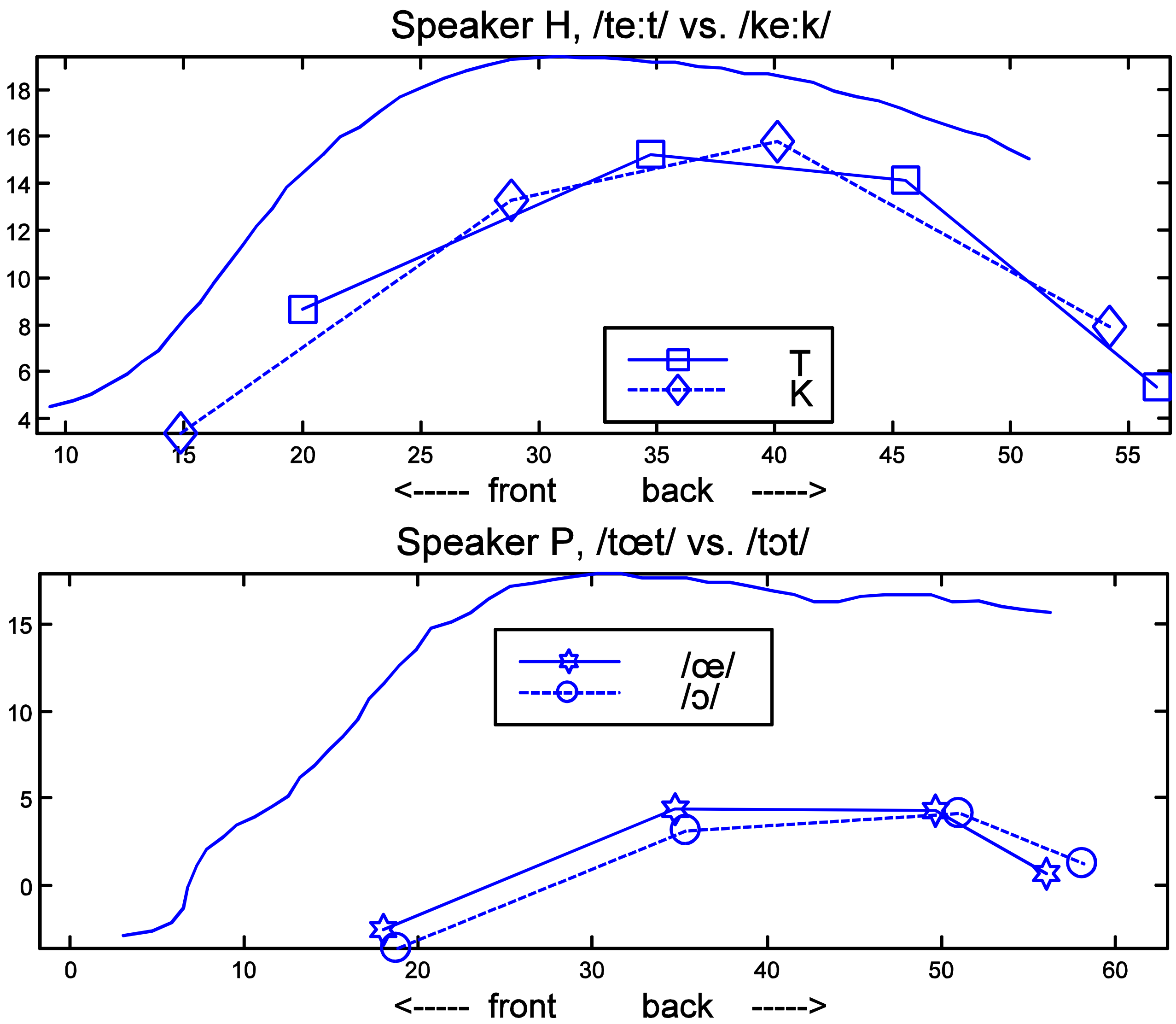




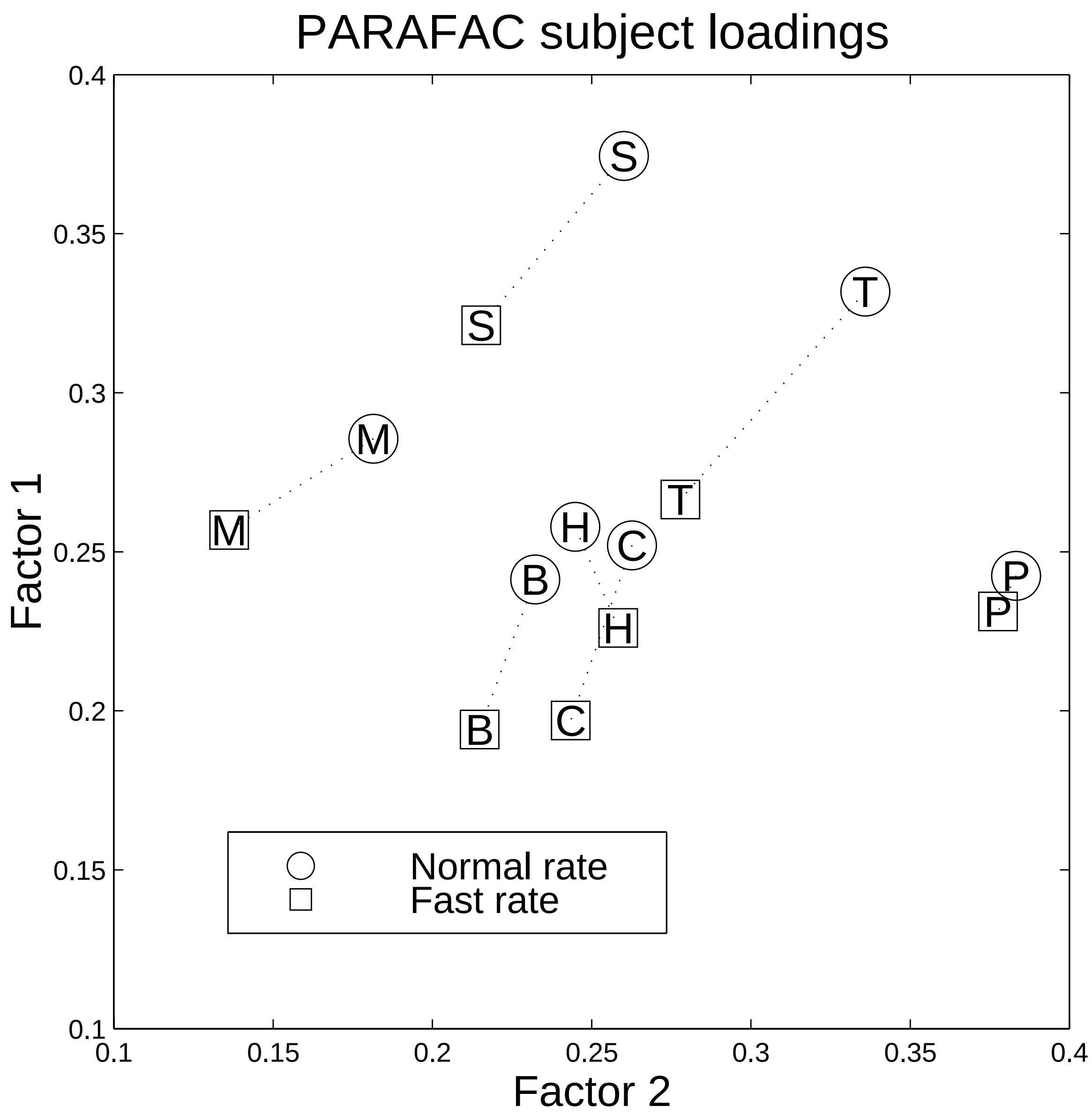
P−context
Tense
Lax
Factor 1
Factor 2
60
40
20
0
−20
−40
−60
−80
−40
−30
−20
−10
0
10
20
30
40
50
60
i
e
y
i
ε
e
ø
y
ø
u
a
u
a
o
o


T−context
Tense
Lax
Factor 1
Factor 2
60
40
20
0
−20
−40
−60
−80
−40
−30
−20
−10
0
10
20
30
40
50
60
i
e
y
i
y
ø
e
ε
ø
u
u
a
o
a
o


K−context
Tense
Lax
Factor 1
Factor 2
60
40
20
0
−20
−40
−60
−80
−40
−30
−20
−10
0
10
20
30
40
50
60
i
e
i
y
y
e
ø
ε
ø
a
u
o
u
a
o



Speaker H, /te:t/ vs. /ke:k/
18
16
14
12
10
8
6
4
10
15
20
25
30
35
40
45
50
55
T
K
<----- front back ----->
Speaker P, /tœt/ vs. /tɔt/
15
10
5
0
0
10
20
30
40
50
60
/œ/
/ɔ/
<----- front back ----->


PARAFAC subject loadings
Factor 1
Factor 2
0.1
0.15
0.2
0.25
0.3
0.35
0.4
S
T
M
H
C
B
P
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?

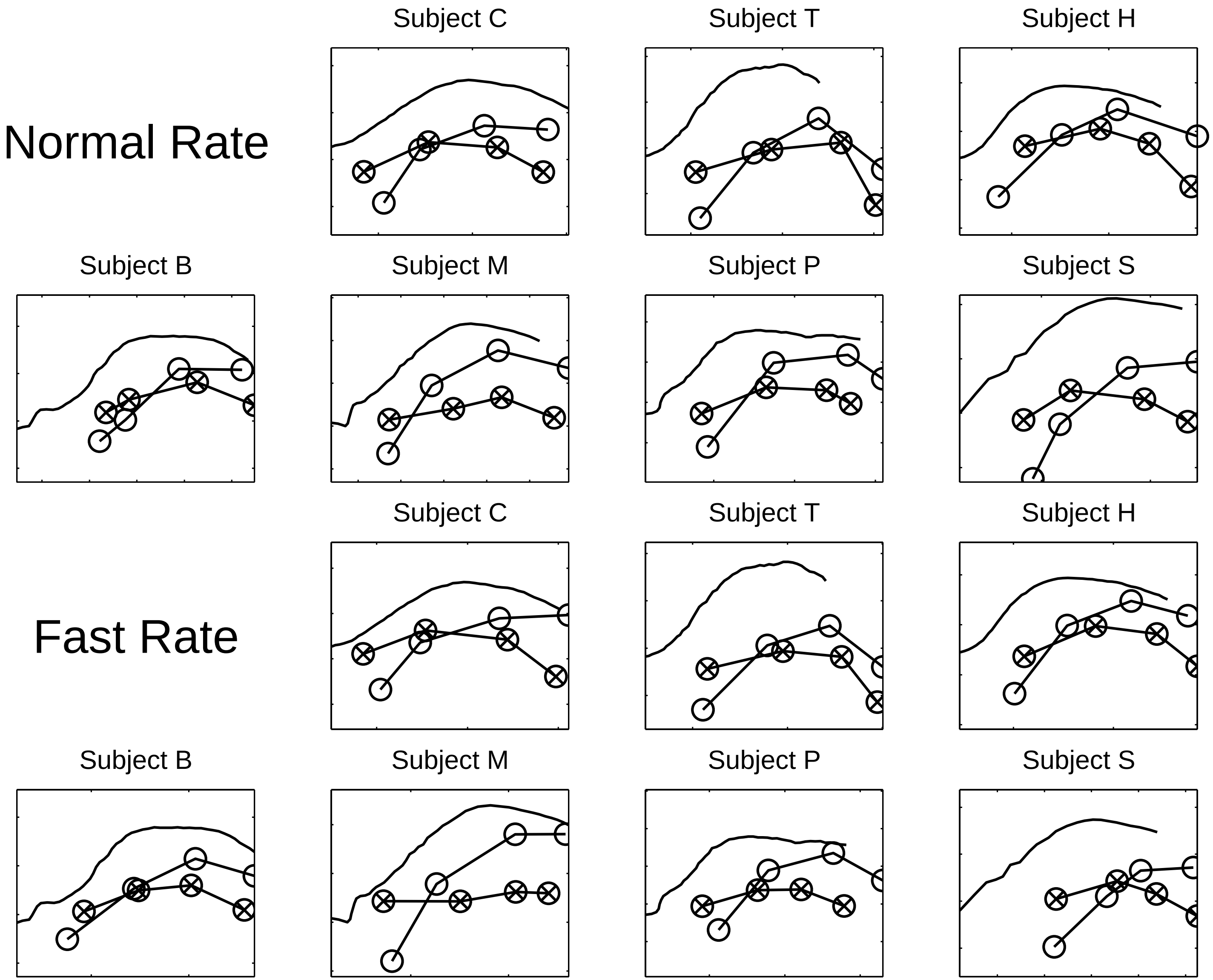
Procedure:

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


Normal Rate
Subject C
Subject T
Subject H
Subject B
Subject M
Subject P
Subject S
Fast Rate
Subject C
Subject T
Subject H
Subject B
Subject M
Subject P
Subject S

Boring brute-force approach

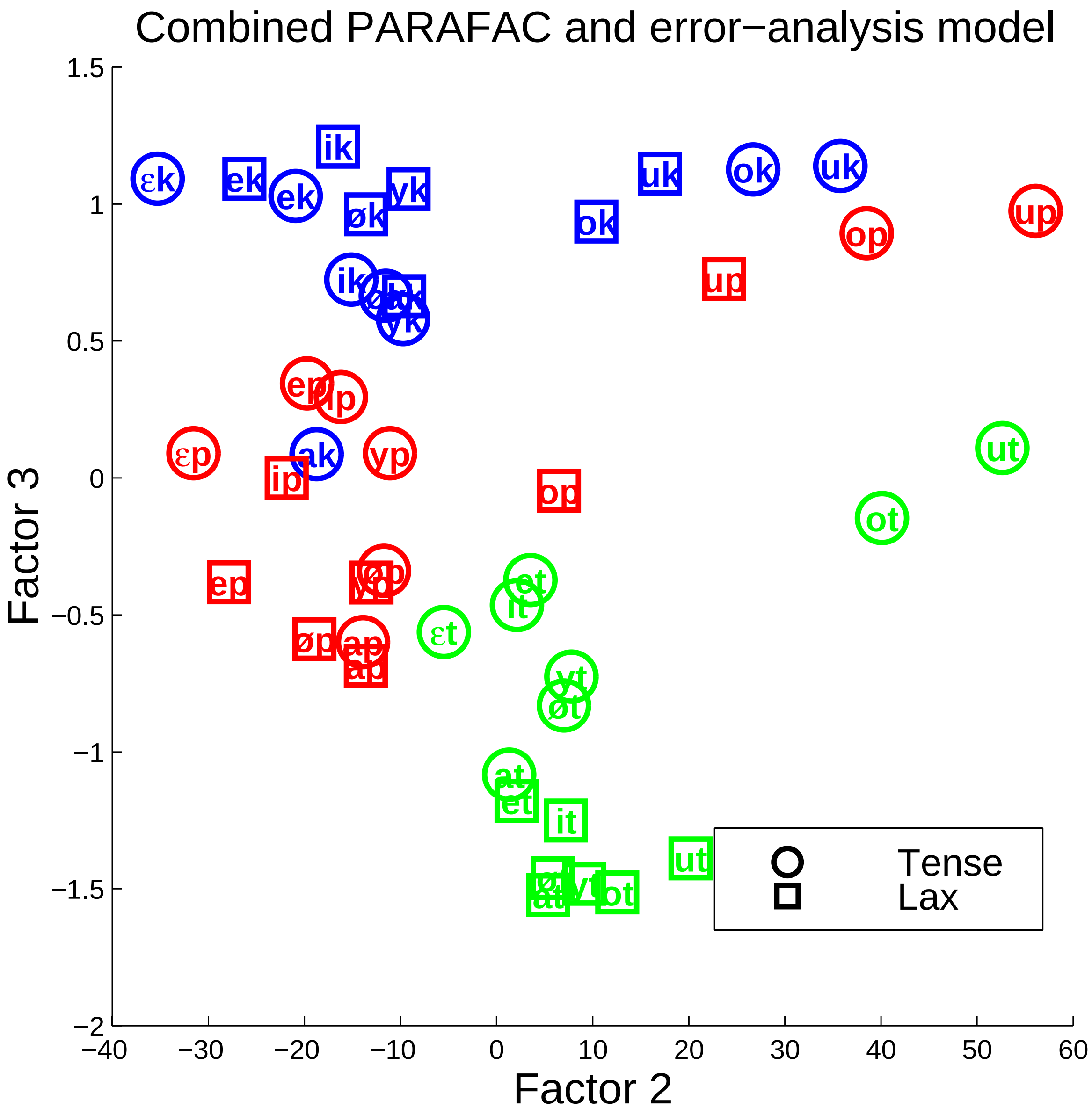
Simply retain for each subject the 45 vowel weights (15 vowels * 3 contexts) and the 8 articulator weights

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



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