ANALYSIS OF TONGUE CONFIGURATION IN MULTI-SPEAKER, MULTI-VOLUME MRI DATA

Phil Hoole[1], Axel Wismüller[2], Gerda Leinsinger[2], Christian Kroos[1,3], Anja Geumann[1], Michiko Inoue[1]

[1] Phonetics Institute, Ludwig-Maximilians-Universtität, Munich
[2] Radiology Department, Klinik Innenstadt, Ludwig-Maximilians-Universtität, Munich
[3] ATR HIP, Kyoto.

email: hoole@phonetik.uni-muenchen.de

Immediate Aims

- What are the nature and the number of the underlying patterns of tongue shapes used in speech (here: vowels)?
- In earlier work (Hoole, 1999a) we derived a two-factor PARAFAC model of tongue shapes for vowels based on EMMA data. Would these results reproduce in analysis based on MRI data?

  Two steps:
  1. Analyze MRI-based midsagittal contours for comparison with this earlier work.
  2. As a potential second stage: Would full 3D data be amenable to the PARAFAC approach?

Subsidiary Aims

(based on same data but not discussed further here)

- Interpretation of patterns of articulatory variability observed in EMMA experiments in terms of area-function variation (Geumann et al., 1999; Hoole, 1999b)
- Reference material for first experiments with 3D EMA system (cf. Zierdt et al., this meeting)

Material and Procedures

- ! 9 speakers
- ! 7 long German vowels /i, e, y, ø, a, o, u/
  (7 of the 9 speakers also recorded the consonants /t, s, n, l, /; these are not discussed further here)
- ! Complete sagittal, coronal and axial scans.
- ! Pixel resolution 1.17mm
- ! Slice thickness 4mm; interslice interval 5mm (4mm for recent sagittal scans)

See grids illustrating slice positions in Figs. 1 and 2

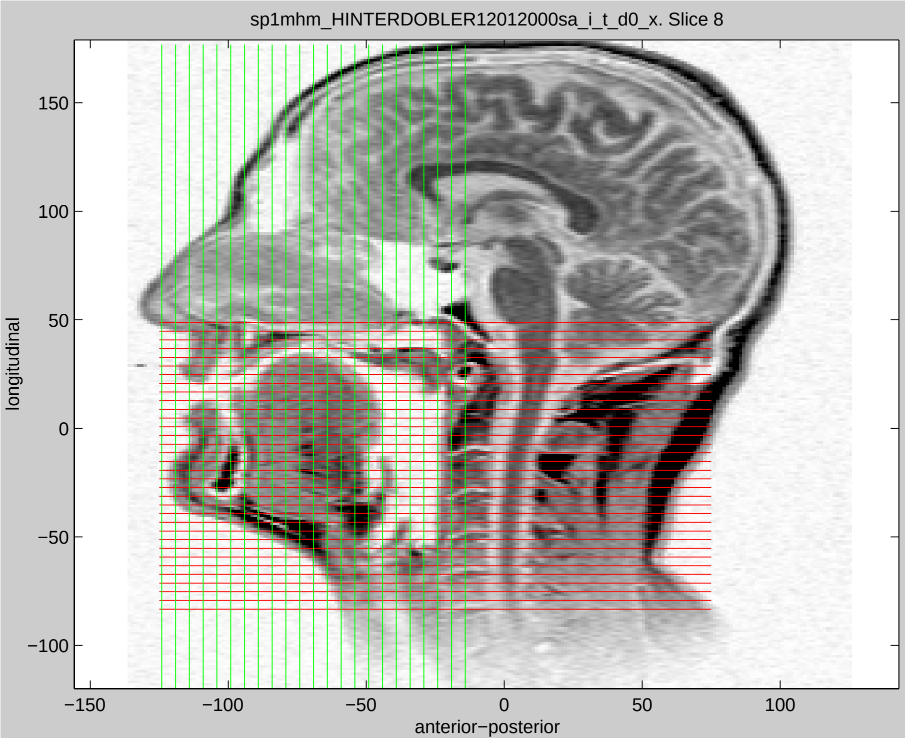

Fig. 1: Coronal and axial slice locations shown on sagittal image

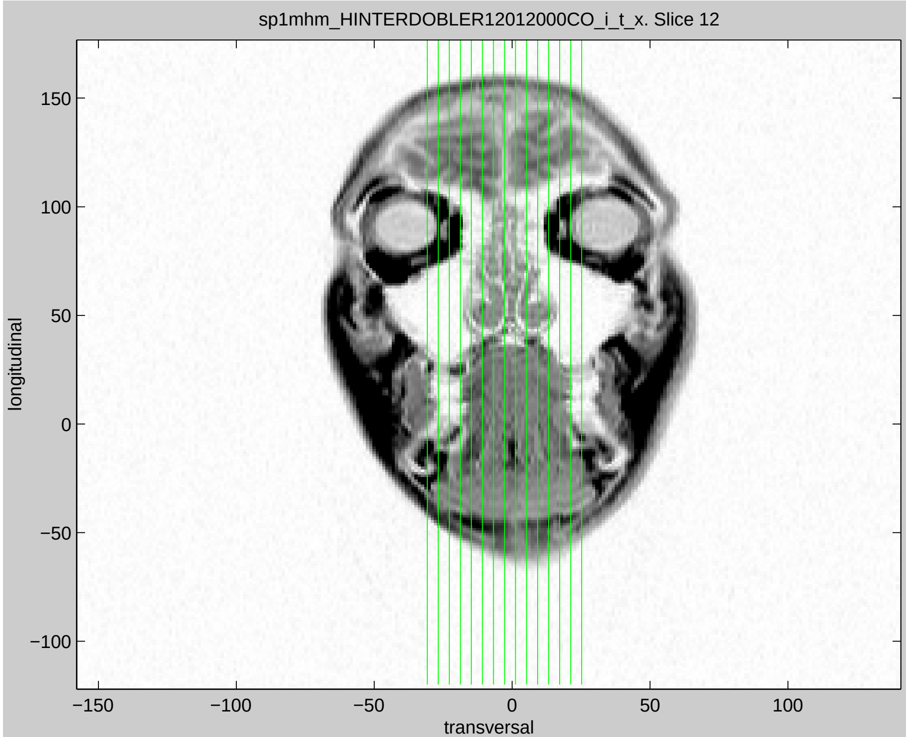

Fig. 2: Sagittal slice locations on coronal image (for earlier subjects more slices were acquired, with correspondingly increased lateral range)

Analysis

The basic set of analyses presented here is confined to midline tongue contours. Nevertheless, the midline contours were not simply taken from the sagittal volumes, but incorporated information from the coronal and axial volumes.
Fig. 3 shows a typical midsagittal tracing obtained in this way.

The contour data for input to the PARAFAC algorithm consisted of the x/y coordinates of 13 “virtual pellets” (from division of each contour into 12 equal-length segments; cf. Nix et al., 1996). See Fig. 4 for examples.

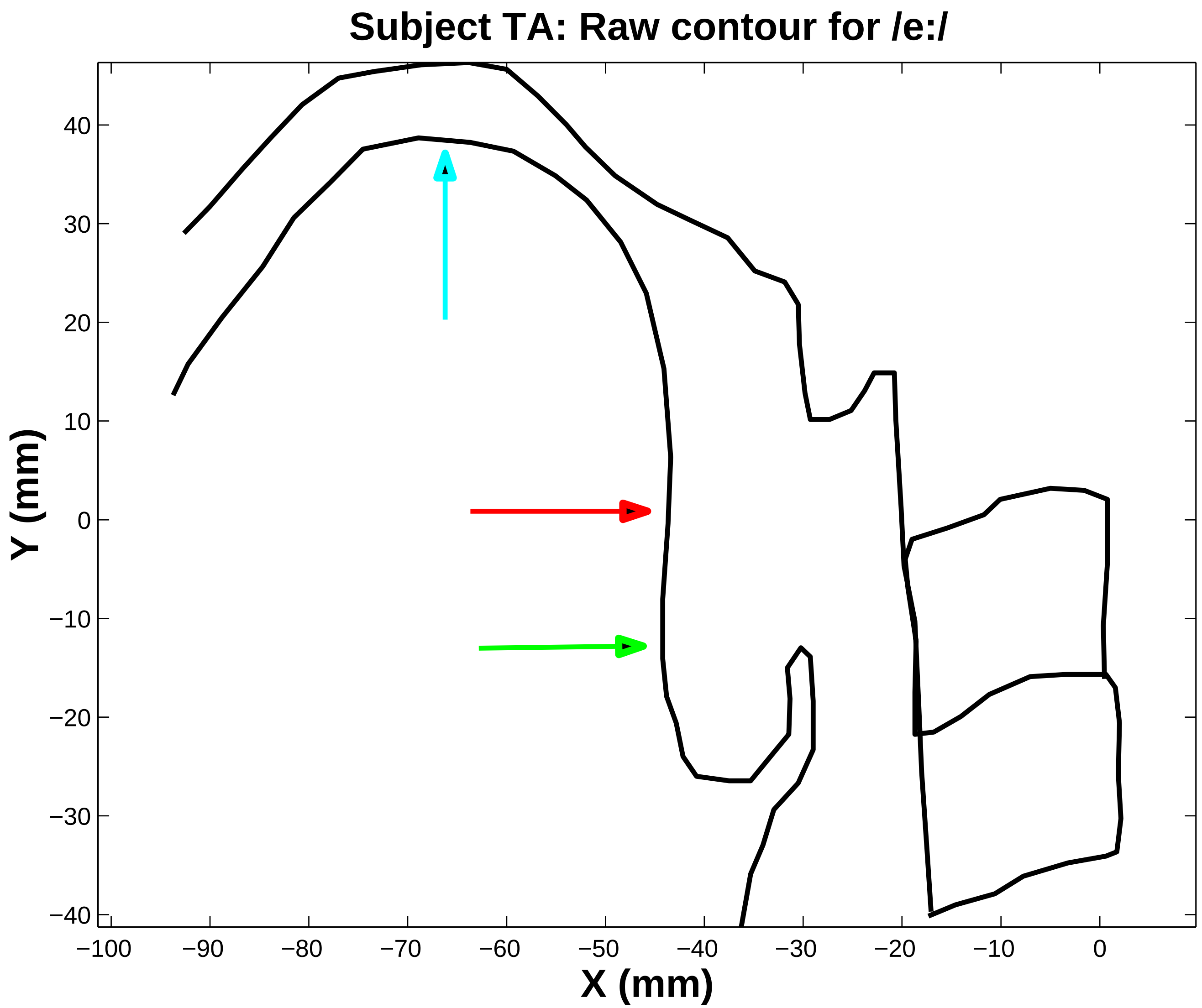

Fig. 3: Examples of raw midsagittal tracings. Green arrow shows height at which tongue contours are truncated (cf. fig.4). Red arrow and cyan arrow indicate approximate location of axial and coronal tongue sections shown in figs. 9 and 10 respectively.

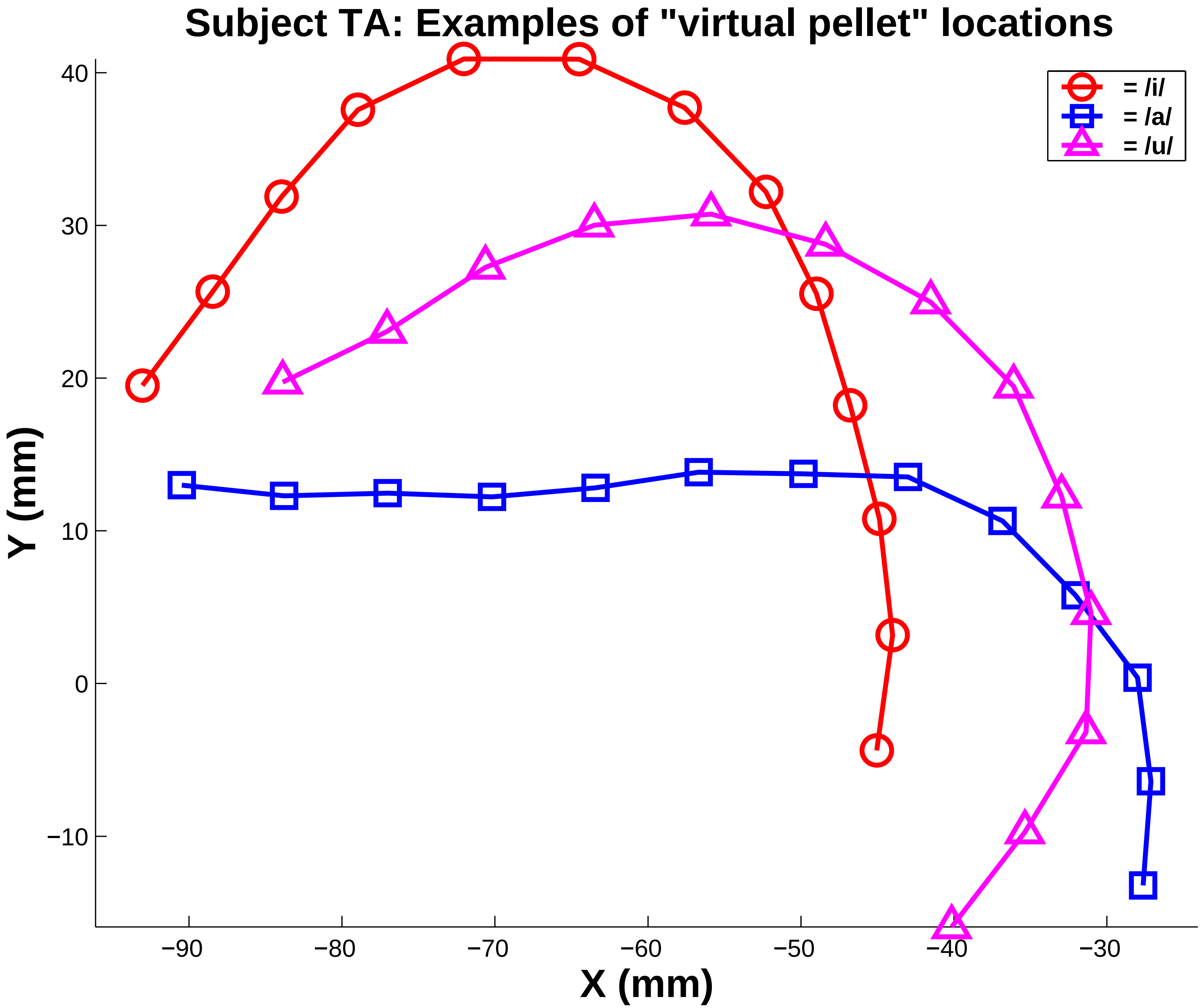

Fig. 4: Examples of the tongue contours for the 3 vowels /i/, /u/ and /a/ given by 13 equidistant points on each contour

Three-Mode Analysis (PARAFAC)

(e.g Harshman et al., 1977)

Systematic exploitation of a third dimension to solve the problem of rotational indeterminacy in the factor axes. The speakers represent this third dimension here.

Model prediction for speaker *k*:

$$\mathbf{Y}_k=\mathbf{A}\mathbf{S}_k\mathbf{V}^T$$

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

The PARAFAC model of the present data

A two-factor model was (eventually, see below) successfully extracted, explaining about 90% of the variance.

The families of tongue shapes associated with each factor are shown in Figs. 5 and 6. The distribution of the German vowels in the factor space is shown in Fig. 7.

Comparison with the PARAFAC model based on EMMA data (Hoole, 1999a):

Factor 1: Here shows more pronounced longitudinal compression.
In Hoole (1999a) all pellets moved forward as they moved up.

Factor 2: Basically similar, but here tongue bunches as it moves back (note also that the overall more retracted tongue position shows more advanced tongue root; this effect would not have been directly observable in the EMMA data).

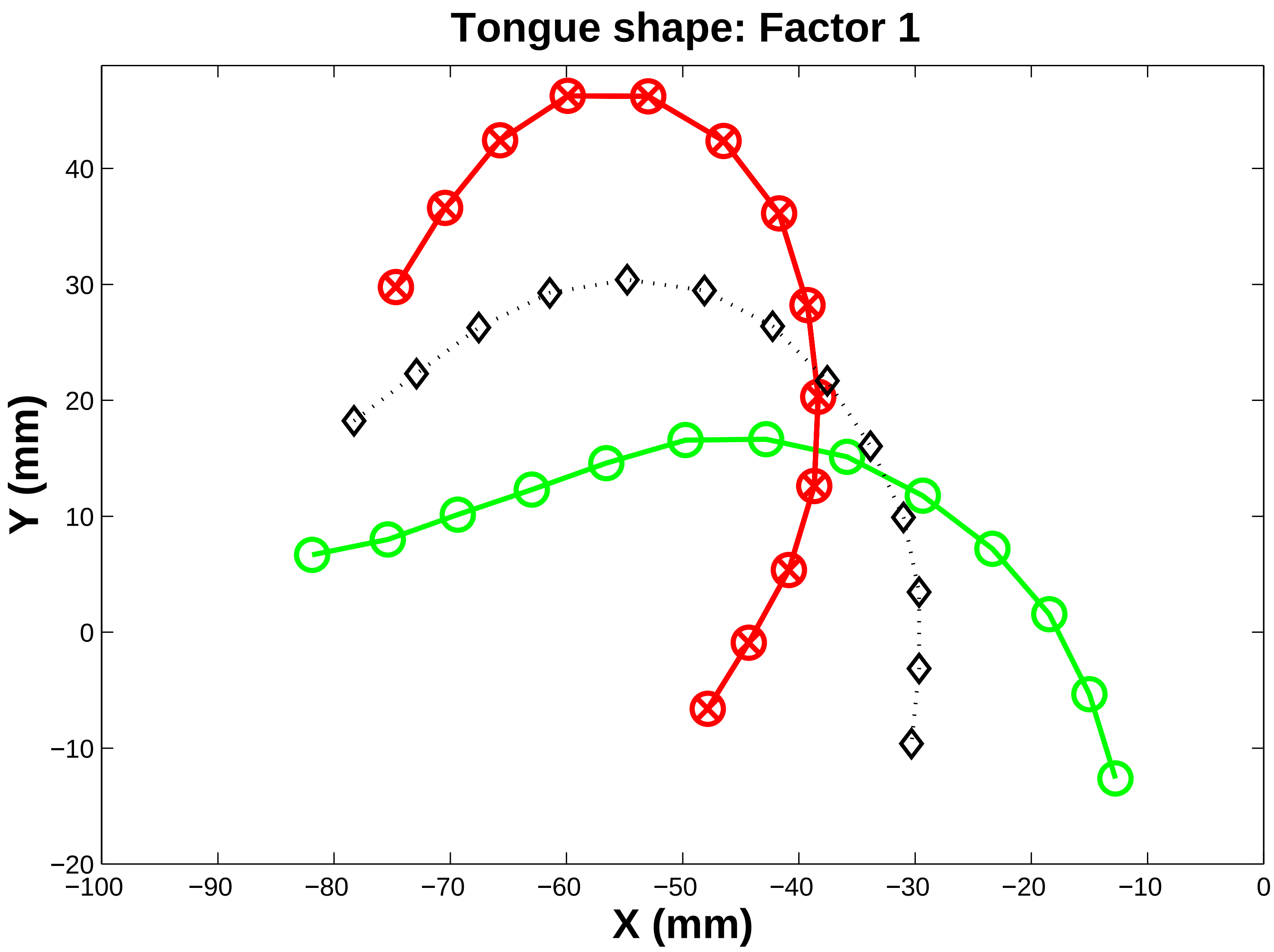

Fig. 5: Factor 1 of the two-factor PARAFAC solution. Deviations from mean tongue configuration (black dotted line with diamonds) obtained by setting the factor to +/- 2 s.d. Positive deviation in green (unfilled circles). Data orientation as in Figs. 1, 3 and 4.

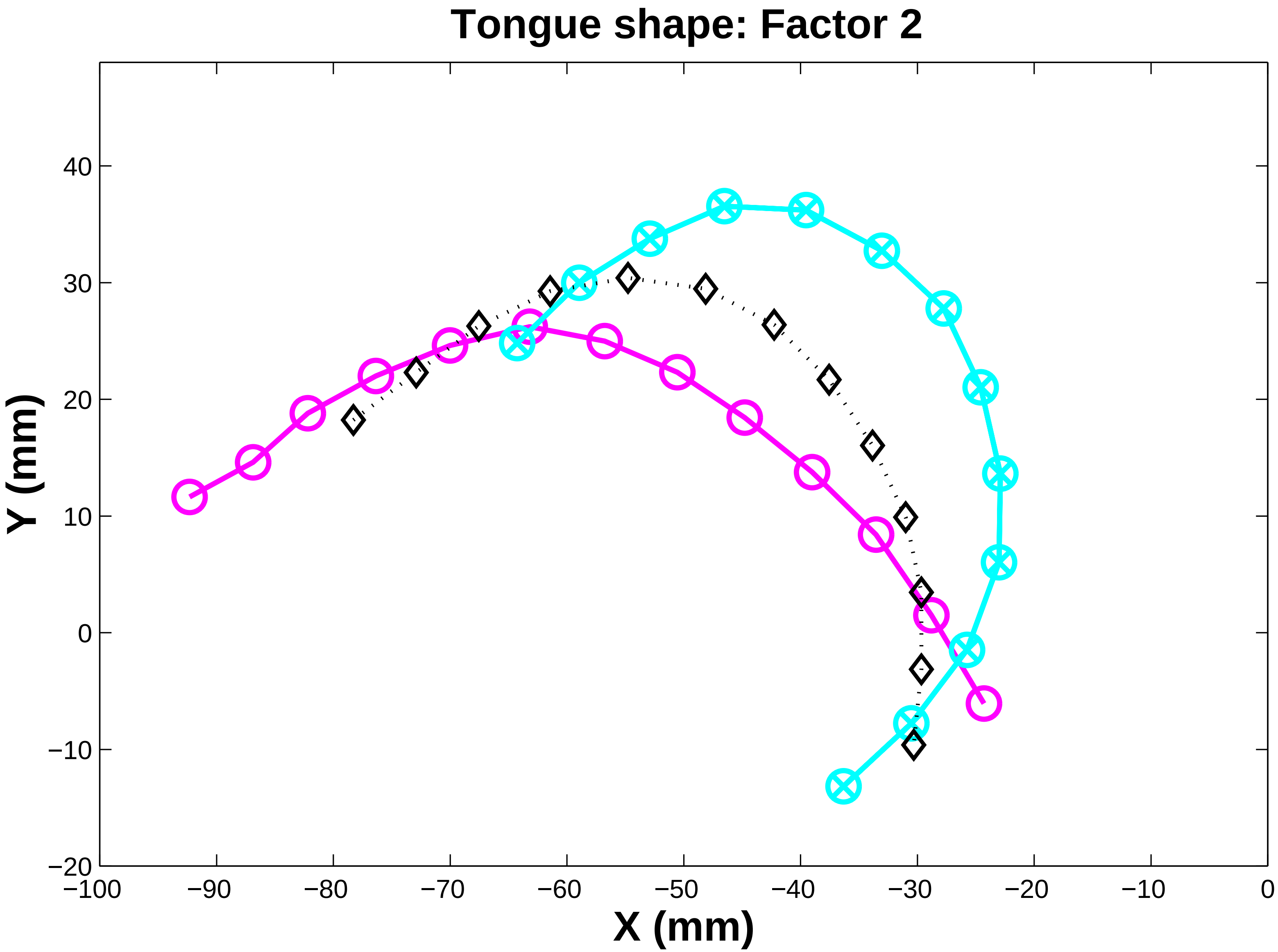

Fig. 6: Factor 2 of the two-factor PARAFAC solution. Other details as in Fig. 5 (positive deviations in magenta (unfilled circles).

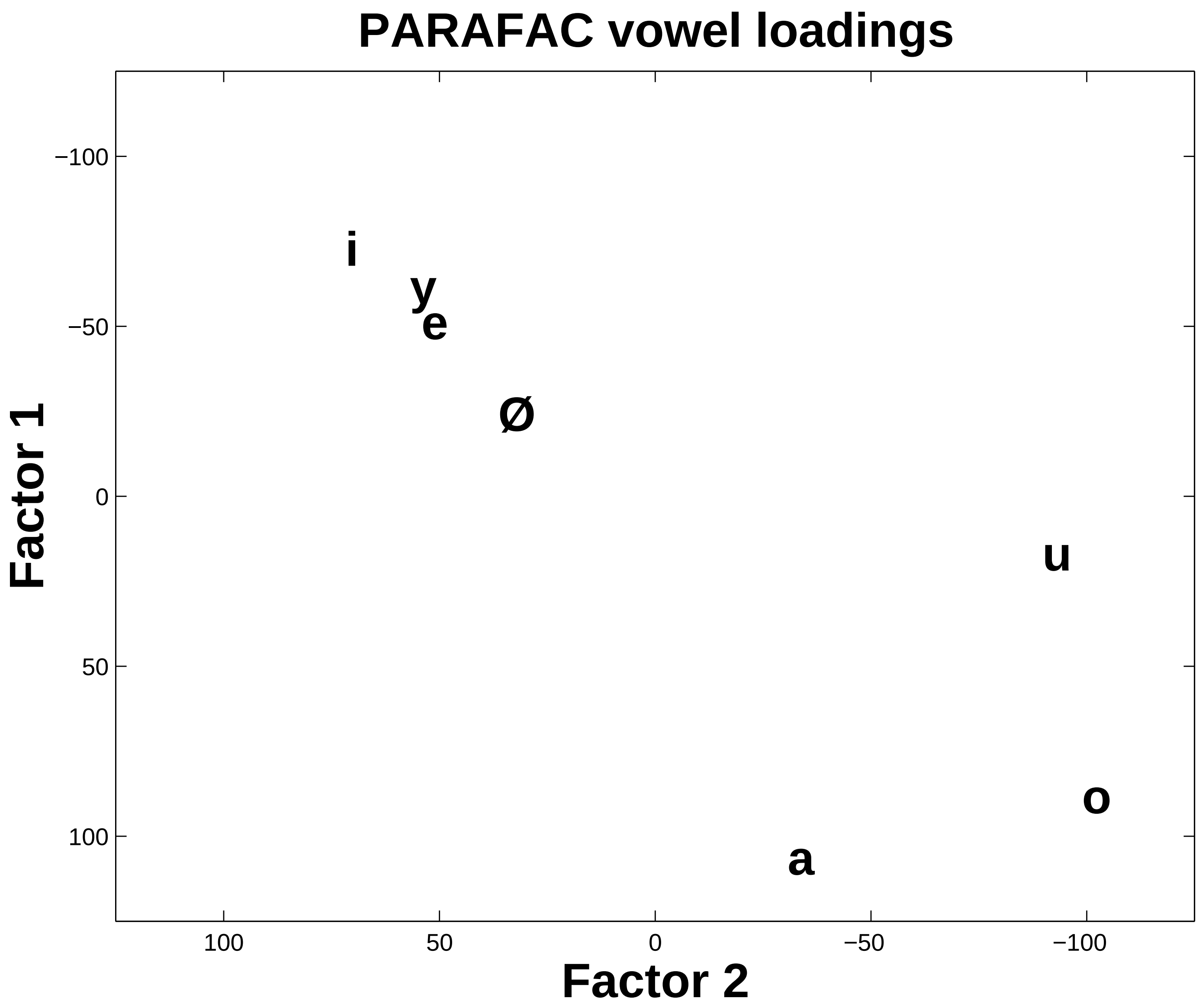

Fig. 7: Distribution of the recorded German vowels in the factor space. (Both axes have been reversed to give a more “traditional” orientation.)

Problem: What is an adequate speaker sample?

The PARAFAC algorithm was first run after 8 of the 9 speakers had been recorded.
This failed miserably to provide a stable model.
But things worked fine when the last speaker, MH, was included in the dataset.
The speaker weights for the two factors are shown in Fig. 8.
Notice that Speaker MH introduces a relatively novel combination of weights.
A range of combinations of speaker weights is necessary to derive the benefit of this three-mode statistical technique.

But how can we know that any given group of speakers will provide an acceptable range of combinations?

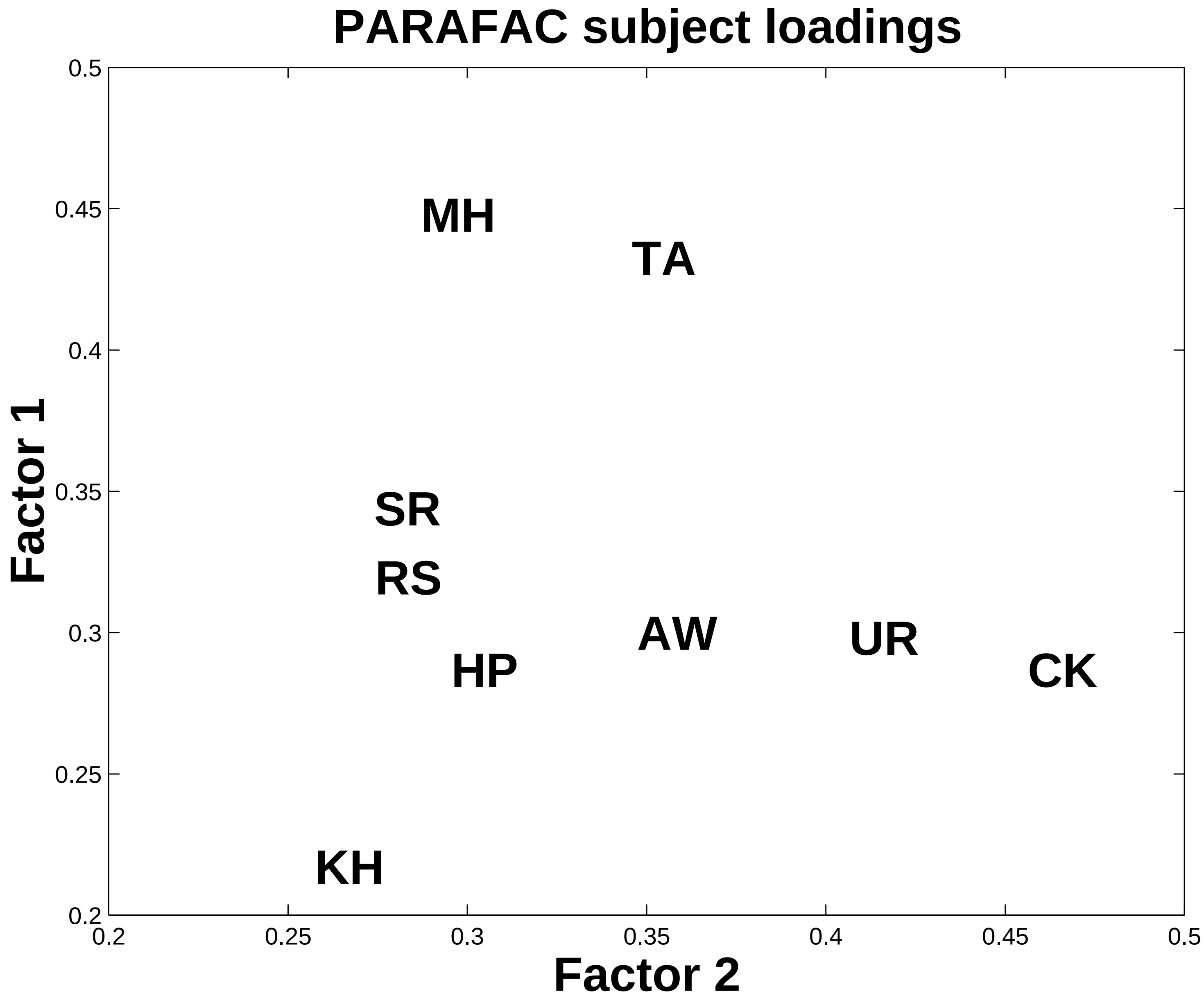

Fig. 8: Subject loadings of the two-factor PARAFAC solution

Preliminary conclusions

! Completely identical solutions do not simply “fall out” when PARAFAC analysis is applied to different kinds of raw data.
Even though PARAFAC has been claimed to be capable of uncovering underlying organizational principles it does nevertheless appear to be sensitive to the precise nature of the input data.

! Although PARAFAC has been in use for many years little is known about what should constitute an adequate speaker sample.

The Next Steps

Can PARAFAC be applied to three-dimensional data?

Possible approach: Use additional parameters to capture the cross-sectional shape of the tongue at each “virtual pellet” position (cf. Stone and Lele, 1992).

Consider the illustrative examples in Figs. 9 and 10.

These show speaker-specific features that would probably not be compatible with the very restrictive PARAFAC model of the mapping between underlying factors and fleshpoint behaviour of individual speakers.

➜ the less restrictive so-called PARAFAC2 model (cf. Geng et al., this meeting)?

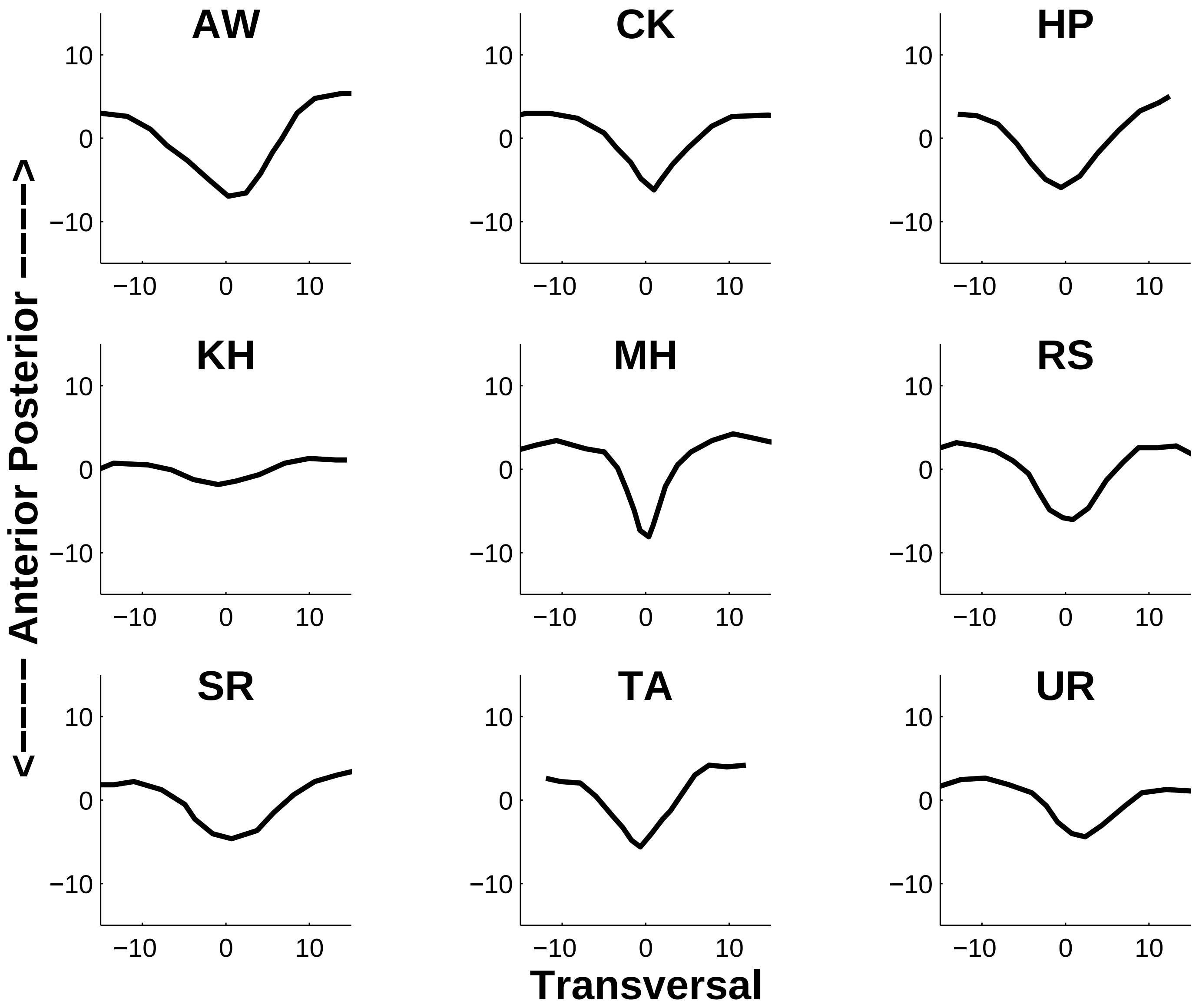

*Fig. 9: Examples of tongue contours for /i/ from **axial** sections. All 9 subjects. Section taken midway between tip of epiglottis and uvula; cf Red arrow in fig.3*

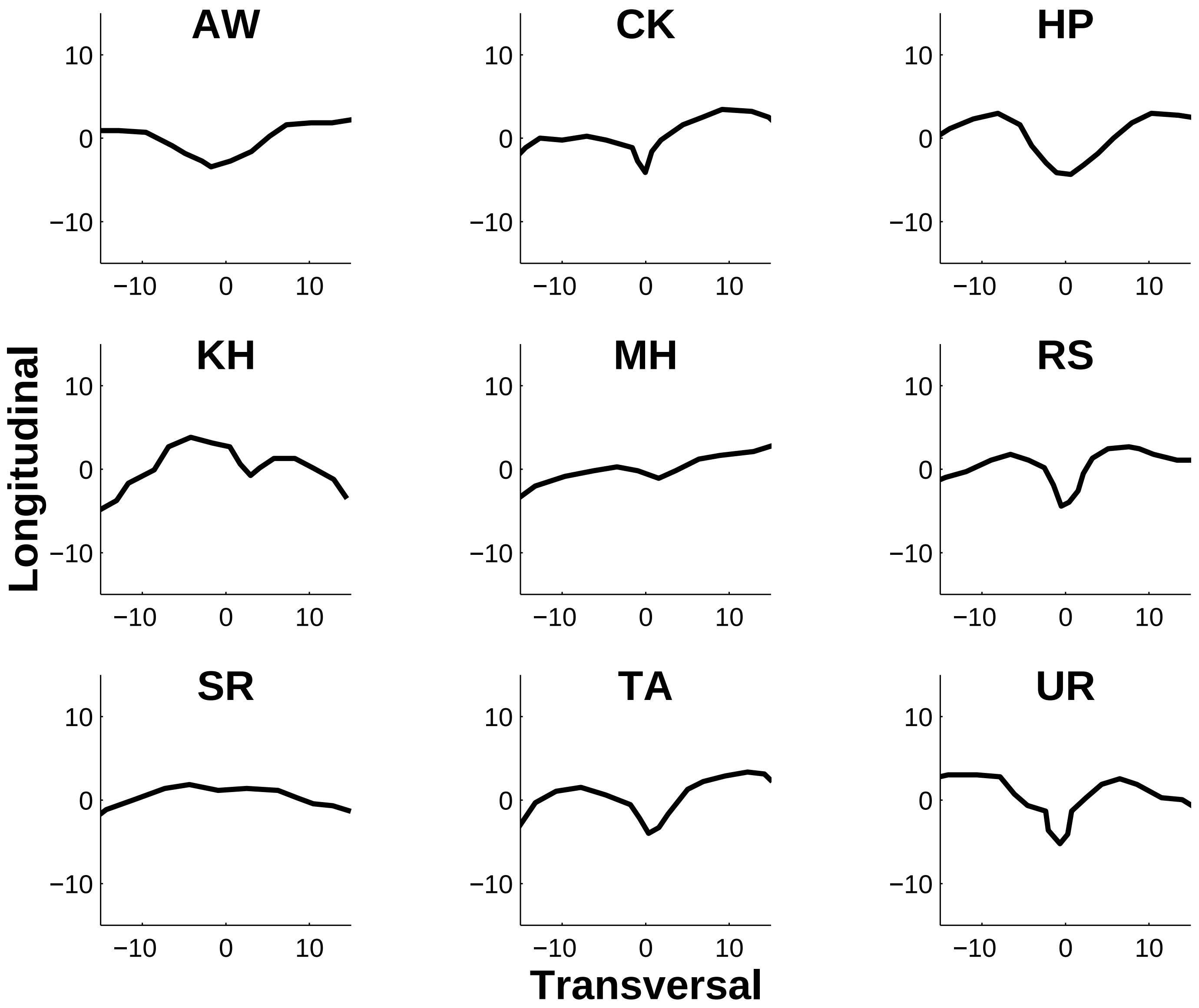

Fig. 10: Examples of tongue contours for /a/ from **coronal** sections. All 9 subjects. Section taken below highest point of palatal vault; cf. cyan arrow in fig. 3

References

Geng, C. & Mooshammer, C. (2000). “Modeling German stress distinction”|. Proc. 5th Speech Production Seminar (this volume).

Geumann, A., Kroos, C. & Tillmann, H.G. (1999). "Are there compensatory effects in natural speech?". Proc. XIVth Int. Cong. Phonetic Sci., pp. 399-402.

Harshman, R., Ladefoged, P., and Goldstein, L. (1977). "Factor Analysis of Tongue Shapes," J. Acoust. Soc. Am. 62, 693- 707.

Hoole, P. (1999a) "On the lingual organization of the German vowel system". J. Acoust. Soc. Am. 106(2), 1020-1032.

Hoole, P. (1999b). "Articulatory-acoustic relations in German vowels". Proc. XIVth Int. Cong. Phonetic Sci., pp. 2153-2156.

Nix, D. A., Papçun, G., Hogden, J. & Zlokarnik, I. (1996). “Two cross-linguistic factors underlying tongue shapes for vowels,” J. Acoust. Soc. Am. 99, 3707-3718.

Stone, M. & Lele, S. (1992). “Representing the tongue surface with curve fits”. Proc. ICSLP 92, pp. 875-878.

Stone, M. & Lundberg, A. (1996). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. 99, 3728-3737.

Zierdt, A.; Hoole, P.; Tillmann, H.G. (1999). "Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements". Proc. XIVth Int. Cong. Phon. Sci., pp. 73- 76.