MODELLING TONGUE CONFIGURATION
IN GERMAN VOWEL PRODUCTION

Philip Hoole

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 Munich, Germany
email: hoole@phonetik.uni-muenchen.de

ABSTRACT

The PARAFAC method of factor analysis was used to investigate patterns of tongue shaping in a corpus of 15 German vowels spoken in 3 consonant contexts by 7 speakers at 2 speech rates, using data from electromagnetic articulography. A two-factor model was extracted, giving a succinct, speaker-independent characterization of the German vowel space and of some important coarticulatory effects on vowel articulation. Moreover, the factors appeared to have a plausible physiological substrate. The PARAFAC model places strong constraints on the form that speaker-specific effects can take, since speaker differences must be captured in a single multiplicative weight per speaker and factor. While these constraints appeared acceptable for modelling vocalic aspects of articulation, more consonantally-related aspects, such as coarticulatory behaviour of the tongue-tip, appeared much more difficult to capture in the PARAFAC framework.

1. INTRODUCTION

A fundamental task in phonetic research is to arrive at a better understanding of how the set of contrasts required by a particular linguistic system on the one hand are implemented by the speech motor system on the other. The linguistic system with which we will be concerned here is the German vowel system, which certainly involves a rich set of contrasts. In this paper we explore a data-driven procedure for deriving a model of vowel articulation. This approach seems justified given that no complete consensus exists for the most appropriate articulatory characterization of vowels. For the central technique to be used, the PARAFAC method of factor analysis, it has been claimed that it can uncover structures in the data that are not just convenient statistical constructs but actually have explanatory power (see references below). Our question essentially boils down to determining how many dimensions underly the tongue shapes that can be observed for vowel articulation, and what their nature is. Direct measurements of the possible physiological building blocks of speech are, of course, very difficult to make, even with EMG (but see Maeda & Honda, 1994); nonetheless, measurements made on the tongue surface should systematically reflect these building blocks, and we may suspect that, due to the limited deformability of the tongue, their number is substantially less than the 8 raw articulatory variables we have available in our data set (corresponding to 2 spatial dimensions measured at 4 sensor locations; see below).

The PARAFAC approach has given phonetically interesting results in several investigations (e.g Harshman, Ladefoged & Goldstein, 1977; Jackson, 1988; Nix, Papcun, Hogden & Zlokarnik, 1996). However, it emerges below that a departure from the PARAFAC framework is necessary to arrrive at a complete model of our data. These extensions to the analysis framework will be covered in a more detailed publication.

PARAFAC requires an inherently three-dimensional data structure, with the third dimension being represented in our case by the speakers. The advantage of PARAFAC over standard two-mode procedures is that it allows the problem of rotational indeterminacy in the orientation of the factor axes to be resolved, giving, it is claimed, greater explanatory power to the factors. At the same time it remains a simple linear model: Given measurements for *nv* vowels from *na* articulators for *ns* speakers, and assuming *nf* factors are extracted, then the results of the PARAFAC procedure are contained in 3 loading matrices **V**, **A** and **S** (for vowels, articulators and speakers) with dimensions *nv*nf*, *na*nf* and *ns*nf* respectively. For speaker *k* the complete dataset $\mathbf{Y}_k$ (dimension *na*nv*) predicted by the model is then given by

$$\mathbf{Y}_k=\mathbf{A}*\mathbf{S}_k*\mathbf{V}^{\mathrm{T}} \qquad (1)$$

where $\mathbf{S}_k$ is a matrix with the *k*th row of **S** on the main diagonal and zero elsewhere, and $\mathbf{V}^{\mathrm{T}}$ is the transpose of **V**.

There are two sides to the simplicity of the model. On the one hand, it is very attractive that speaker-specific and speaker-independent effects are explicitly separated; on the other hand, the model makes very strong assumptions about the form that these speaker-specific effects can take, i.e each factor is simply scaled by a single speaker-specific weight for all vowels. Attempting to apply the PARAFAC algorithm to a dataset can thus provide insight into the extent to which these assumptions are justified for human speech behaviour.

2. MATERIAL AND PROCEDURE

All 15 monophthongal vowels of German (/iː,ɪ;yː,ʏ;eː,ɛ;ɛː;øː,œ;ɑː,a;oː,ɔ;uː,ʊ/) were spoken in 3 symmetrical consonant contexts (/p, t, k/), embedded in the carrier phrase *“Ich habe geCVCe gesagt”*. Five repetitions of each CVC combination were produced by 7 speakers (6 male, 1 female) at two speaking rates (normal, fast), recorded in separate sessions.

Articulatory movements were monitored by means of electromagnetic midsagittal articulography (EMMA, Carstens AG100). In this paper the x/y cooordinates of 4 sensors mounted on the tongue will be analyzed (details of experimental procedure in Hoole, 1996). For each vowel, one frame of articulatory data was extracted at the midpoint of the vowel. The data was then averaged over the five repetitions of each CVC combination.

3. DEVELOPING THE PARAFAC MODEL

3.1 A False Start

For PARAFAC analysis it is necessary to choose the number of factors on which to base the model.This was assessed by applying a (two-mode) principal component analysis to each speaker individually. If the speaker-specific differences are consistent with the PARAFAC model then PARAFAC should be able to model the complete data set using the number of factors typically appropriate for individual speakers in a PC-analysis. In these analyses three factors consistently accounted for the data well. It thus appeared warranted by the data, and phonetically plausible, to base the PARAFAC model on three factors. This was unsuccessful. The algorithm failed to converge. This suggests that some aspects of the structure of the dataset are inconsistent with the PARAFAC model, and suspicion falls most obviously on the influence of consonantal context, as this represents the most substantial extension of our dataset compared with earlier, successful applications of the PARAFAC model. We will return briefly to this below.

Before attempting further analysis of the complete dataset it now appeared necessary to analyze the dataset separately for each consonantal context, firstly in order simply to confirm that our data is amenable to analysis under conditions comparable to other reported investigations, and secondly in order to provide a baseline against which to judge further attempts at getting to grips with the full data set.

3.2 Models For Individual Consonant Contexts

We present first the results for the vowels spoken in the /p/-context, as this can be regarded as the most neutral consonantal context with regard to lingual articulation. One would expect a two-factor solution to be appropriate for a dataset involving only one consonantal context. This indeed turned out to be the case. The two-factor solution was clearly reliable. The unexplained variance amounted to 7.7% and the RMS error to 1.24 mm. This is very much par for the course. For the /k/-context a two-factor model was also successfully extracted; the model was very similar to the /p/-context one.

Surprisingly, the extraction of a two-factor model for the /t/-context ran into problems. The algorithm took longer to converge than in the /p/ and /k/ contexts and the resulting solution gave strong signs of being degenerate. Moreover the solution was substantially different from the /p/-case.

A degenerate solution may be caused by inconsistency of the data with the PARAFAC model. As we discuss in more detail elsewhere (Hoole, in preparation), it turns out that the way tongue blade raising is captured by the front two EMA sensors exhibits speaker-specific patterns that are inconsistent with the simple multiplicative PARAFAC model. And clearly this problem is most relevant in the /t/-context.

These separate analyses of individual consonant contexts had indicated what the ideal result for a complete model might be (i.e an RMS modelling error in the region of 1.2 mm) and also enabled potential problems in the data to be localized. The aim was now to proceed back towards a model for the complete data set.

3.3 Models For Multiple Consonantal Contexts

As a first step back we tested whether a successful two-factor model could be extracted when the data involving the two 'easy' consonant contexts /p/ and /k/ were analyzed together. This proved to be the case. Compared with /p/ and /k/ taken individually, the unexplained variance deteriorated somewhat, but the model for combined /p/ and /k/ was very similar to the models extracted for /p/ and /k/ separately.

Since this step had been successful we then restored the /t/-context material to the dataset and extracted a *two*-factor solution for the complete dataset. This was also successful in the sense that the algorithm converged readily to a reproducible solution, and no evidence of degeneracy was found. Not surprisingly, however, there was a further noticeable increase in model error, unexplained variance now amounting to 20% and the RMS error to 1.9 mm.

In Hoole (in preparation) we show how the PARAFAC model error can be further analyzed to extract an additional articulatory component essentially capturing the coarticulatorily-induced alternation between tongue-blade and tongue-dorsum raising for the /t/ vs. /k/ contexts.For the purpose of the present paper we will concentrate on discussing the two-factor PARAFAC solution just extracted from the complete dataset. It seems justifiable to use this as our basic model of vowel articulation since the two-factor solution extracted from the complete dataset is still very similar to the solutions for the simple 'p-only' or 'k-only' data.

4. DISCUSSION OF THE PARAFAC MODEL

4.1 Tongue Configurations

The weights for all 8 articulator coordinates with respect to each factor are plotted as a pattern of tongue displacement around average tongue position using averaged speaker weights. The result is shown in the two panels of Fig. 1.

The first factor shape looks quite similar to the first factor derived by Harshman et al., and referred to by them as 'front raising'. In our Fig. 1 we see substantial raising (and some advancement) of the front part of the tongue, and advancement (with some raising) of the rear part of the tongue. Whether our factor 2 is as closely related to the Factor 2 of Harshman et al. (referred to by them as 'back raising') is less clear. Our Factor 2 shows some raising at the rear coil locations but the retraction component is more marked.

4.2 The Vowel Space

The three panels of Fig. 2 show, separately for each consonant context, how the German vowel system is represented in the space of the first two factors.

Factor 1 has been allotted to the ordinate since it has the strongest tongue-raising component; however, since neither factor exclusively involves raising vs. lowering, or advancement vs. retraction, the vowel space mapped out by the two factors is rotated with respect to traditional phonetic representations of the vowel space. The extreme vowels for each factor are /i:/ and /o:/ for Factor 1 and /ɛ:/ and /u:/ for Factor 2.

Let us first discuss features of the vowel space that are similar over consonant context, before turning to some important differences.

We will look first at the contrast between tense and lax vowels. Here we need to consider front and back vowels separately. We find for the front vowels and /a/ that the lax variant takes on less extreme values (i.e closer to zero) for Factor 1. However, a consistent pattern with respect to Factor 2 is not discernible. For the back vowels /u/ and /o/ the situation is different since it is now Factor 2 rather than Factor 1 that shows the more consistent pattern: Lax vowels show less extreme values with respect to Factor 2.

Comparing front unrounded and rounded vowels it is clearly the case that the rounded cognates occupy less extreme positions with respect to Factor 1. In fact, every front rounded vowel is actually closer on the Factor 1 dimension not to its direct unrounded cognate, but to the phonologically next lowest unrounded vowel (/y:/ closer to /e:/ than to /i:/, etc.). However, the unrounded-rounded contrast also involves slightly but consistently more negative values of Factor 2 for unrounded (i.e these show, roughly speaking, more fronting than the rounded vowels).

Let us now consider differences in the vowel space for the different consonantal contexts. Perhaps the most striking feature is the distribution of the vowels with respect to Factor 2 for /t/-context compared to the other two contexts. In /t/-context essentially all vowels except the tense back vowels /u:/ and /o:/ cluster close to zero; the range of variation along the factor 2 dimension is compressed, compared to the other two contexts. This probably provides part of the reason why we encountered difficulties in extracting a stable 2-factor solution for /t/-context vowels on their own. Considering Factor 2 primarily as an advancement-retraction dimension, the effect is thus essentially one of retraction of the front vowels (and /a/) in /t/-context. This is so substantial that there is no overlap in Factor 2 values for front vowels in /t/-context with their values in the other two contexts. Note also that the nominally front vowel /œ/ is located very close to the back vowel /ʊ/ in this context. It should be remarked that these strong coarticulatory effects captured by Factor 2 involve tongue-body placement (clearly, the tongue-tip is also relevant, but cannot be considered here; see Hoole, in preparation). In fact, as far as we are aware, this very simple yet basic finding that front vowels in /t/-context have a more retracted tongue-body position than in /k/-context has not yet been reported. Although it may seem counterintuitive at first blush, it is probably a natural strategy to provide the tongue-tip with room to elevate to form the alveolar closure.

A final, briefer observation related to coarticulatory effects remains to be made. The most neutral context /p/ shows very clearly an effect that has provoked much debate for almost a century (Fischer-Jørgensen, 1985), namely that /ɪ/, the lax cognate of /i:/ is substantially lower (here in terms of Factor 1) than the next lowest tense vowel /e:/ (ceteris paribus for /y:/). However, when coarticulatory effects are taken into account this effect becomes blurred: In /k/-context /ɪ/ has about the same value as /e:/, and /ʏ/ is somewhat higher than /øː/. Again this is probably an easily explainable effect: /k/-context tends to elevate tongue-body position and does so relatively more for the lax vowels.

5. GENERAL DISCUSSION

Evidence has now accumulated from a number of investigations that control of tongue configuration for speech is organized around a small number of underlying components. The present study provides strong confirmation for this, on the basis of a particularly large data set. Moreover, there are grounds for thinking that the two factors extracted in our PARAFAC analysis could have a plausible physiological substrate. Specifically, Factor 1 could well reflect the agonist-antagonist pairing of Genioglossus Posterior and Hyoglossus proposed by Maeda & Honda (1994), while Factor 2 could reflect the Genioglossus Anterior and Styloglossus pairing suggested by them.

The PARAFAC approach places strong constraints on the form that speaker-specific effects can take, thus allowing an extemely parsimonious representation of multi-speaker data sets. This approach has proved quite successful in investigations of vowel articulation. However, the present study indicates that more consonantally-related aspects (in our case coarticulatory activity of the tongue tip) may not be compatible with this framework. In a forthcoming publication we will indicate how the PARAFAC model can be supplemented by speaker-specific principal component analysis of the PARAFAC model error to generate a hybrid model that also gives a systematic account of these consonantal effects, but still remains quite parsimonious.

ACKNOWLEDGEMENTS

Work supported by German Research Council Grant TI -69/29-3. For this work we used a MATLAB package described in Bro, 1997 (available online from http://newton.foodsci.kvl.dk/). We are grateful to M. Tiede for drawing our attention to this package.

REFERENCES

Bro. R. (1997). "PARAFAC. Tutorial and applications". *Chemom. Intell. Lab. Syst*., 38(2), 149ff.

Fischer-Jørgensen, E. (1985) "Some basic vowel features, their articulatory correlates, and their explanatory power in phonology", in V. Fromkin (ed.) *Phonetic Linguistics, Essays in honour of Peter Ladefoged*, pp. 79-99.

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

Hoole, P. (1996). "Issues in the acquisition, processing, reduction and parameterization of articulographic data", *FIPKM*, 34, 158-173.

Hoole, P. (in preparation). “Modelling tongue position in German vowels”.

Jackson, M. T. T. (1988). “Analysis of tongue positions: Language-specific and cross-linguistic models”, *J. Acoust. Soc. Am*. 84, 124-143.

Maeda, S. & Honda, K. (1994). "From EMG to formant patterns of vowels: The implication of vowel spaces". *Phonetica*, 51, 17-29.

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

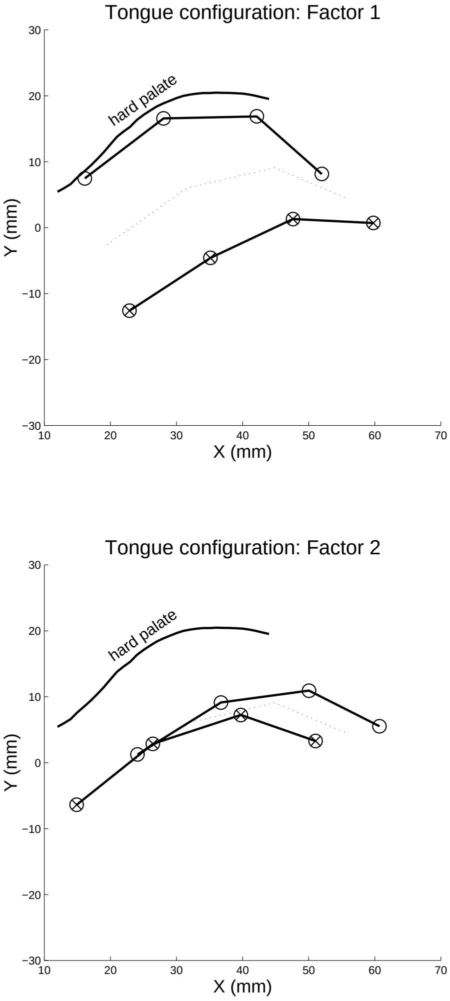

Fig. 1: *Tongue shapes related to the factors of the two-factor PARAFAC model of the complete dataset. Each panel shows displacement using mean speaker weights from mean tongue position (shown by dotted line) caused by setting each factor in turn to +/- 2 standard deviations (positive deviation: empty circles; negative deviation: circles with crosses). More anterior locations are to the left. Palate contour is an average of overlapping portions of the palate contours of the seven speakers.*

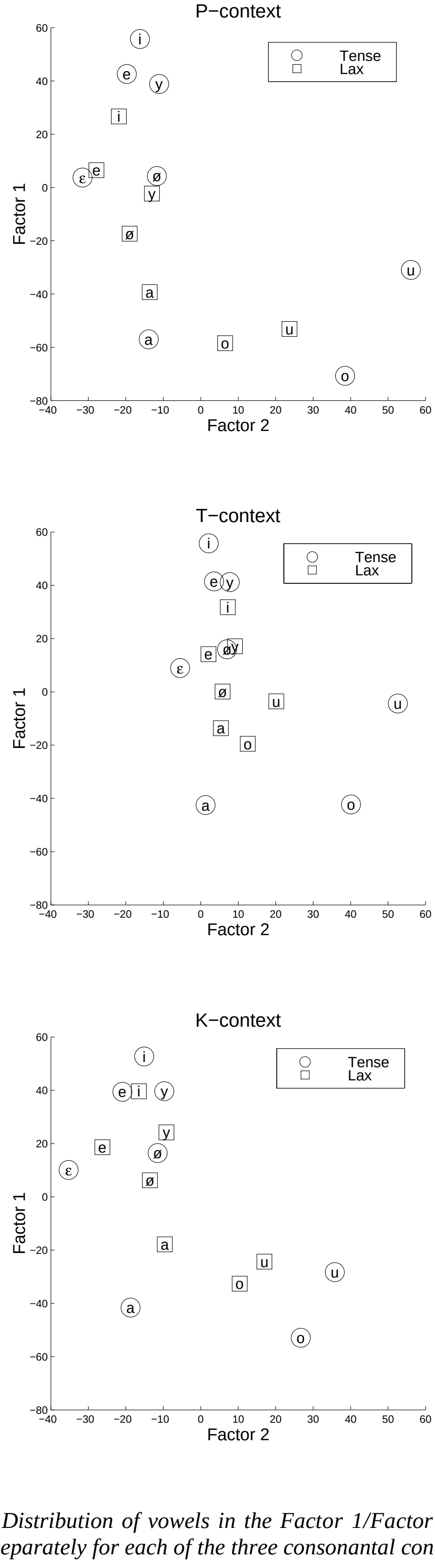

Fig. 2: *Distribution of vowels in the Factor 1/Factor 2 space, shown separately for each of the three consonantal contexts.*