# MODELING THE GERMAN STRESS DISTINCTION

Christian Geng and Christine Mooshammer
*ZAS -Research Centre for General Linguistics, Berlin, Germany*
geng@zas.gwz-berlin.de, timo@zas.gwz-berlin.de

**ABSTRACT**

Low dimensional and speaker-independent linear vocal tract parametrizations can be obtained using the 3-mode PARAFAC factor analysis procedure first introduced by Harshman et al. (1977). The following study used PARAFAC to investigate the stress distinction in German vowel production. Tongue movements of six German speakers were recorded by means of EMMA. The speech material consisted of the 15 German vowels, recorded in /t/- context. Our corpus includes these vowels in stressed and unstressed position. They were entered into the classical PARAFAC1 model treating the stress distinction for each subject as two different speakers. This gave a reasonable 2-factor solution, but was not without drawbacks. The model turned out to be capable of recovering gross anatomical properties of our subjects, but failed to return intraindividual differences in tongue shapes with respect to word stress. This indicated that the strict linearity assumptions required in the classical PARAFAC model were too strong to capture stress-specific variation in full detail. We supposed that a model closely related to PARAFAC, PARAFAC2, should allow to account for systematic variation produced by word stress by imposing weaker structure on the data. As will be shown, PARAFAC2 modeled the physical properties of the vocal tract shape in a more realistic and plausible way.

## 1 INTRODUCTION

One broad research area aiming at a deeper understanding of the motor implementation of linguistic contrasts has been the search for efficient characterizations of vocal tract shapes by factor analytic methods (e. g. Jackson, 1988, Maeda, 1990). In factor analytic tradition, the PARAFAC approach has given phonetically interesting results in a variety of studies and thus is quite well evaluated. Different topics have been under consideration following Harshman's et al. prototypical study. Examples are crosslinguistic issues concerning the number of factors in different languages (Jackson, 1988, Nix, 1996) or paralinguistic features as speech rate (Hoole, 1999).

PARAFAC is a type of multi-mode analysis procedure and therefore contrasting with Principal Component Analysis (PCA) or factor analysis, which are two mode representations. PARAFAC requires an at least three-dimensional data structure with the third dimension usually being represented by different speakers, i.e. if all speaker weights are fixed to be one, then PARAFAC reduces to PCA. The advantage of PARAFAC is that there is no rotational indeterminacy as in PCA, in other words, PARAFAC gives unique results. The PARAFAC (in accordance with literature from now on called PARAFAC1) model can be written as (following Kiers et al., 1999, alternative notations are given in Harshman et al., 1977 or Nix et al., 1996)

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

where $\mathbf{X}_k$ is the $k$th slab of the input data matrix, with $k$ the number of speakers, **A** is the matrix of articulators, **S** is the matrix for speakers and **V** the matrix for vowels. The matrix of articulator weights is held constant for each slab of the data cube, i. e. for all $k$ speakers. This addresses Cattell's notion of parallel proportional profiles:

"The basic assumption is that, if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study." (Cattell and Cattell, 1955, citation Harshman and Lundy, 1984, p. 151). Another way to put it is this (Harshman 1977, p. 609): "Thus if speaker A uses more of Factor 1 than does speaker B for a particular vowel, then speaker A must use more of factor 1 than speaker B in all other vowels. The ratio of any two speakers' usage of a given factor must be the same for all vowels."[1]

This does not always have to be a plausible assumption though; it can also turn out to be too restrictive in some cases. For illustration, the other extreme would be to put no structure at all onto **A** –which is equal to reducing the PARAFAC model to a PCA and loosing the desirable uniqueness properties. Between the two extremes of having all $\mathbf{A}_k$ equal to **A** and having $\mathbf{A}_k$ unconstrained there are several possibilities for imposing structure in $\mathbf{A}_k$. PARAFAC2 offers one such intermediate possibility:

$$\mathbf{X}_k=\mathbf{A}_k\mathbf{S}_k\mathbf{V}^T \qquad (2)$$

Within PARAFAC2, each loading matrix $\mathbf{A}_k$ is modeled as $\mathbf{A}_k=\mathbf{P}_k\mathbf{A}$, $k$=1,...,$K$, where $\mathbf{P}_k$ is an $I$*$R$ column orthonormal matrix and **A** is of size $R$*$R$, with $R$ denoting the number of factors. Note that **A** is of different dimensionality here than it is in the PARAFAC1 model. The diagonal matrix **A** represents the common part of the speaker-specific matrices **P** in an $R$-dimensional subspace (Kiers et al., 1999). This investigation wants to focus on potential advantages of using a relaxed model like PARAFAC2.

## 2 EXPERIMENTAL PROCEDURE

Six native speakers of German (4 males, JD, PJ, CG and DF and 2 females, SF and CM) were recorded by means of an electromagnetic midsagittal articulographic device. The speech material consisted of words containing /tVt/ syllables with nuclei (V= /i,ɪ,y,ʏ,e,ɛ,ɛː,ø,œ,aː,a,u,ʊ,o,ɔ/) in stressed and unstressed positions. Stress alternations were fixed by morphologically conditioned word stress and contrastive stress. So each symmetric CVC sequence was embedded in the carrier phrase *Ich habe tVte, nicht tVtal gesagt. (I said _, not _)* with the first test syllable /tVt/ always stressed and the second always unstressed. For each of the 15 vowels, between six and ten repetitions of these vowels were recorded. Tongue, lower lip and

[1] Note: Speaker A in the preceding citation has nothing to do with **A**, the matrix of articulators.

jaw movements were monitored by EMMA (AG100, Carstens Medizinelektronik). Four sensors were attached to the tongue, one to the lower incisors and one to the upper lip. Two sensors on the nasion and the upper incisors served as reference coils to compensate for helmet movements during the recording session. The analyses in this study are limited to the four transducers on the tongue. Simultaneously, the speech signal was recorded by a DAT recorder. Articulatory movements were smoothed by a lowpass filter at 30 Hz and articulatory data extracted at a vowel specific tongue configuration. The data were then averaged over the six respectively ten repetitions of each vowel. This strategy is different compared to radiographic studies, e.g. the Harshman et al. study, where no repetitions of items are possible.

## 3 DEVELOPMENT OF THE MODEL

One central goal of statistic modeling of the vocal tract shape is the derivation of parsimonious representation. Therefore it is revealing to have a look at the relationship between the number of parameters estimated by such a model and a constant number of points in the data matrix. Our data have following dimensions: 15 German vowels * 8 articulators (four x/y positions in two dimensions) * 6 speakers (for reasons of transparency, the accentuation distinction is neglected in the next paragraph).

As the number of robust factors over several PARAFAC studies has consistently been two, we decided a priori to fit 2-factor solutions exclusively. The number of parameters to estimate in a PARAFAC1 model is given by F(I + J + K) in the case of a trilinear model, with I, J, K denoting the number of vowels, articulators and speakers respectively. A 2-factor PARAFAC1 solution with I=15 vowels, J=8 (four x and 4 y) coordinates and K=6 speakers has 2*(15+8+6)=58 parameters to estimate. The other extreme, the equivalent 2-mode 2-factor PCA model with the person mode collapsed has to estimate 2*(L+J) parameters, with L being here of the dimension I*K (I and K as defined above) In a 2-factor PCA this amounts to 2*(90+8)=196 parameters. Between these two extremes, the PARAFAC2 model takes an intermediate position and has F(I+K*J+J)=138 parameters to estimate.

In a next step, we discuss the alternatives for incorporating the stress distinction in a PARAFAC model. One possibility would be to capture word stress in the vowel mode, i. e. calculate separate vowel weights for stressed and unstressed cognates. The dimensionality of the input data would amount to 30 vowels * 8 x/y positions * 6 speakers. There are two main arguments against this alternative. The *first* argument is a systematic one: This kind of model would tear apart the German vowel system with its 15 vowels; the *second* argument is related to the assumptions that have to be met if fitting is successful. The alternative model, which represents the accentuation contrast in the speaker mode (input dimensions 15 vowels * 8 x/y-positions * 12(=2*6)) speakers requires the 2*6 speaker weights not to vary at random over the stressed/unstressed sessions and therefore a violation of these assumptions can be revealing.

After the decision for this global model structure, the logical next step was to determine the pre-processing strategy. As Harshman and Lundy (1984b) point out, preprocessing can be a matter of trial and error and previous knowledge. With respect to PARAFAC1, it turned out to be viable to pre-process the data as follows: After averaging over individual tokens –as described in B. - the overall mean was determined for each subject and subtracted from the data. The data fed into the subsequent PARAFAC1 algorithm thus consisted of displacements from the average articulatory configuration of each subject.

For fitting the PARAFAC2 model, we had to slightly refine our pre-processing strategy: In contrast to the PARAFAC1 model, double centering across vowels and coordinates turned out to be the best among several workable solution.

Furthermore, for both models, we had to constrain the factors in the vowel mode to be orthogonal, because there was a tendency of the model to collapse. This is a drawback, because a constrained model will fit the data less well than an unconstrained model, but if the constrained model is more interpretable and realistic, this may justify the decrease in fit (Bro, 1998). There are several arguments for hazarding the consequences connected with this procedure. First, fitting a PARAFAC model in the /t/-context ran into problems in another investigation (see Hoole, 1999 for a discussion of possible reasons for /t/-context degeneracies). The second argument addresses issues of generalizability/external validity: A good model should be generalizable over a broad class of situations i.e. consonantal contexts. The third argument is that the chosen PARAFAC1 solutions revealed no signs of degeneracy in terms of the core consistency statistic, a newly developed tool for assessing the fit of a PARAFAC1 model[2].

As a consequence of constraining the model, the common diagnostic tools for identifying degenerate solutions and assessing the reliability of a whole model cannot be used in the traditional way. The triple product of the correlations between corresponding sets of weights for each pair of factors will always be zero. The statistic therefore looses its potential to identify degenerate solutions. A milder version of testing for factor intercorrelations is to show stability over several reruns which succeeded.

Concerning the other common model diagnostics, our results can be summarized as follows: The RMS-error amounted to about 2mm for PARAFAC1 and a little less than 1.40 for the corresponding PARAFAC2 model. The corresponding explained variances were about 96% for PARAFAC1 and 93% for the PARAFAC2 model.

## 4 RESULTS

### 1 Articulator Weights

With regard to the projections of the articulator weights in the 2-factor space, the results can be summarized as follows: The described PARAFAC1 model was capable of capturing the gross anatomic properties of our subjects, but it did not capture the stress distinction within subjects, or, in other words, the subjects' x/y coordinates were basically the same for stressed and unstressed vowels. A possible interpretation for these results lies in the distinction between front and back vowels. Centralization for back vowels –roughly- works in the opposite spatial direction than for front vowels, but PARAFAC1 models the stress distinction by only one set of weights for each person and the resulting tongue shapes can be thought of as a "model-induced compromise": Front and back vowels tend to cancel each other out with respect to determining different speaker weights. In contrast, our PARAFAC2 model revealed interpretable stress differences. In an interindividually consistent fashion we recovered more extreme tongue configurations for the stressed data set. Fig. 1 shows speaker-independent tongue

[2] The core consistency diagnostic should be (approximately) 100%, for a discussion see Bro (1998)

projections, i. e. compares mean stressed speaker weights with mean unstressed weights.
We now turn to an interpretation of our factors. First it is important to note that our PARAFAC2 factors are permuted compared to previous studies. Our first factor bore some

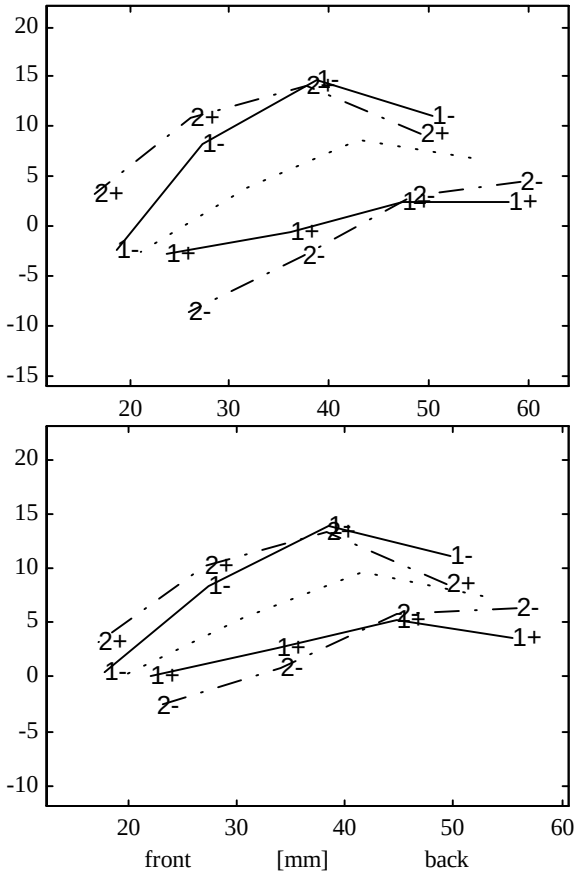


*Fig. 1. Maximum and minimum tongue shapes related to the two-factor PARAFAC2 model. The top panel shows 2 SD displacements using the mean of the stressed subset of speaker weights. The bottom panel is equivalent but is related to unstressed subject weights.*

resemblance with Harshman et al.'s second factor -particularly the constriction in the velar region- which they referred to as „back raising". In turn, our second factor was similar to what these authors referred to as „front raising". In addition to the exchange of factors, there was a sign change concerning our first factor, i. e. higher tongue positions were associated with negative signs. The similarity of our tongue shapes with Harshman's figures is limited due to methodological reasons, as discussed in Nix et al., 1996, p. 3708: „Although measuring the shape of the tongue with respect to anatomically normalized vocal diameter gridlines does reduce the initial representational dimension, this measurement scheme needlessly looses information such as the positions of the tip of the tongue in the horizontal dimension. More importantly, the range of possible solutions is artificially constrained by the orientation of the grid lines. For example, a factor representing protrusion and/or retraction of the tongue tip is not possible because no grid line is oriented in this direction." Thus it is not too surprising that both of our factors contain a quite strong horizontal component, as our data are „fleshpoint data".
Summarizing these results in view of the stress distinction a) in contrast with PARAFAC1, the PARAFAC2 method is capable of capturing the stress distinction with respect to tongue shapes and b) stressed vowels exhibit more extreme articulator configurations than their unstressed counterparts.

## 2 Subject Weights

The pattern of subject weights further legitimates the chosen way to model the data: There are no sign changes between subjects indicating that our factors are not used to capture

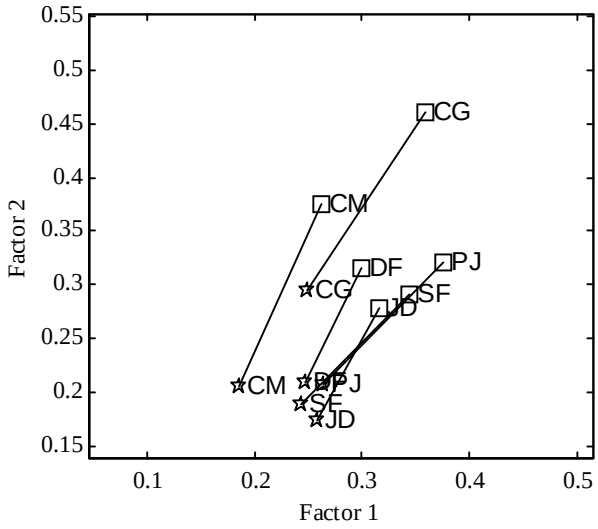


*Fig. 2. Distribution of subject weights in the factor1/factor2 space. Representations of stressed cognates are indicated by squares, unstressed cognates are represented by stars.*

subject-specific features. This is valid for both of our models. The relationship between the weights for stressed vowels and unstressed vowels can be expressed as follows: For all speakers it is the case that stressed vowels have higher factor loadings than unstressed vowels. It is revealing to compare our findings with results obtained by Hoole (1999), as in their study speech rate was assessed following a very similar rationale. As Farnetani (1990, pp. 109f.) points out, there are interindividual differences in the implementation of speech rate: „a) a reduction in the amplitude of the movements with no changes in velocity (...), b) a reduction in the amplitude and an increase in velocity (...), c) no changes in the amplitude and an increase in velocity (...) d) reductions in both amplitude and velocity."
Hoole's speaker weights (1999, Fig. 6, p. 1027) for speech rate indicate an interindividually divergent implementation of speech rate, whereas in our results a more homogeneous shift in the direction of higher loadings for word stress is prevailing. (see Fig. 2)

## 3 Vowel Weights

The traditional representation of the German vowel system was the most problematic part of these analyses: As mentioned, the models could only be fit using orthogonality constraints in the vowel mode. The results indicate that this led to unnatural projections of the German vowel system, especially with respect to the „artificial" second factor: Differences in the horizontal

dimensions get exaggerated due to not reaching the absolute minimum of the error/loss function to be minimized (see Fig. 3) Thus further interpretation of vowel weights is skipped here.

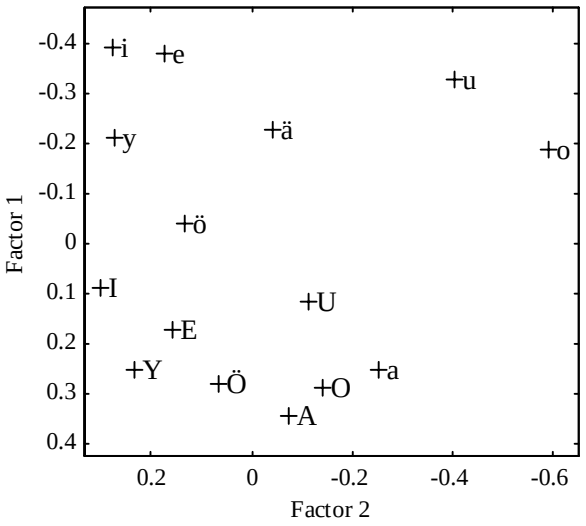


*Fig. 3. Projection of the German vowel system in the factor1/factor2 space. Lower-case letters indicate tense vowels, upper-case letters lax vowels.*

## V. DISCUSSION

This investigation focused on methodological issues related to the German stress distinction.

Methodologically, the well-known PARAFAC1 algorithm was compared with the closely related PARAFAC2 algorithm. It was described as the less restrictive, but less parsimonious variant of PARAFAC1. It was shown that PARAFAC2 has the potential to give revealing results in a situation where PARAFAC1 is too restrictive.

We limited our analysis to two factors, but there is a third replicable factor accounting for the alternation between tongue-blade and tongue-dorsum-raising in two-way factor analytic studies, which has not yet been captured within a three-way model. We believe that this should be possible by means of PARAFAC2 with data more robust against coarticulatory effects than the /t/-context data we were using.

With respect to our „target variable“, word stress, our results further substantiate the finding that the motor implementation of word stress is less speaker-specific than the implementation of speech rate: Its implementation is completely consistent across speakers and therein differs from speech rate, which exhibits speaker-specific patterns. In our three-dimensional design, we could show this simultaneously in two of our interpretational modes, i. e. for speaker weights *and* for articulator weights using PARAFAC2.

The last point concerns modeling error and summarizes our experience with PARAFAC2: PARAFAC2 performs better in terms of modeling error with regard to tongue positions. This is consistent with the model: PARAFAC2 uses more parameters to estimate the data, thus is less parsimonious but can compensate this disadvantage through higher flexibility in certain situations.

## 6 REFERENCES


[1]Bro, R. (**1997**). PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* **38**.

[2]Bro, R. (**1998**). The N-way on-line course on PARAFAC and PLS. http://www.models.kvl.dk/

[3]Bro, R., Andersson, C. A. & Kiers, H. A. L. (**1999**). PARAFAC2-Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics* **13**, pp. 295-309

[4]Cattell, R. B. and Cattell, A. K. S. (**1955**). Factor rotation for proportional profiles: analytical solution and an example. *British Journal of Statistical Psychology* **8**, pp. 83-92.

[5]Farnetani, E. (**1990**). Lingual coordination and its spatiotemporal domain. In *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal. (Kluwer, Dordrecht), pp. 93-130.

[6]Harshman, R., Ladefoged, P & Goldstein, L. (**1977**). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America* **62**, pp. 693-707.

[7]Harshman, R. A. & Lundy (**1984a**). The PARAFAC model for three-way factor analysis and multidimensional scaling. In *Research methods for multimode data analysis*, edited by H. G. Law, C. W. Snyder, J. A. Hattie and R. P. McDonald. (New York: Prager), pp. 122-215.

[8]Harshman, R. A. & Lundy (**1984b**). Data preprocessing and the extended PARAFAC model. In *Research methods for multimode data analysis*, edited by H. G. Law, C. W. Snyder, J. A. Hattie and R. P. McDonald. (New York: Prager), pp. 216-284.

[9]Hoole, P. (**1999**). On the lingual organization of the German vowel system. *Journal of the Acoustical Society of America* **106**, pp. 1020-1032.

[10]Jackson, M. T. T. (**1988**). Analysis of tongue positions: Language-specific and cross-linguistic models. *Journal of the Acoustical Society of America* **84**, pp. 124-143.

[11]Kiers, H. A. L., ten Berge, J. M. F & Bro, R. (**1999**). PARAFAC2-PartI. A direct algorithm for the PARAFAC2 model. *Journal of Chemometrics* **13**, pp. 275-294.

[12]Maeda, S. (**1990**). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 131-149.

[13]Mooshammer, C., Fuchs, S. & Fischer, D. (**1999**). Effects of stress and tenseness on the production of CVC syllables in German. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 409-412

[14]Nix, D. A., Papcun, M. G., Hodgen, J. & Zlokarnik, I. (**1996**). Two cross-linguistic factors underlying tongue shapes for vowels. *Journal of the Acoustical Society of America* **99**, pp. 3707-3718.