Some comments on the reliability of three-index factor analysis models in speech research

Christian Geng
Zentrum für Allgemeine Sprachwissenschaft, Berlin

Phil Hoole
Institut für Phonetik und sprachliche Kommunikation der LMU, München

---

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) and discussed in a series of subsequent papers in the Journal of the Acoustical Society of America (Jackson (1988), Nix et al. (1996), Hoole (1999), Zheng et al. (2003)). Nevertheless, some questions of importance have been left unanswered, e.g. none of the papers using this method has provided a consistent interpretation of the terms usually referred to as "speaker weights". This study attempts an exploration of what influences their reliability as a first step towards their consistent interpretation. With this in mind, we undertook a systematic comparison of the classical PARAFAC1 algorithm with a relaxed version, of it, PARAFAC2. This comparison was carried out on two different corpora acquired by the articulograph, which varied in vowel qualities, consonantal contexts, and the paralinguistic features accent and speech rate. The difference between these statistical approaches can grossly be described as follows: In PARAFAC1, observation units pertain to the same set of variables and the observation units are comparable. In PARAFAC2, observations pertain to the same set of variables, but observation units are *not* comparable. Such a situation can be easily conceived in a situation such as we are describing: The operationalization we took relies on the comparability of fleshpoint data acquired from different speakers, which need not be a good assumption due to influences like sensor placement and morphological conditions.
In particular, the comparison between the two different approaches is carried out by means of so-called "leverages" on different component matrices originating in regression analysis, calculated as $v = diag(A(A'A)^{-1}A')$ and delivering information on how "influential" a particular loading matrix is for the model. This analysis could potentially be carried out component by component, but we confined ourselves to effects on the global factor structure. For vowels, the most influential loadings are those for the tense cognates of non-palatal vowels. For speakers, the most prominent result is the relative absence of effects of the paralinguistic variables. Results generally indicate that there is quite little influence of the model specification (i.e. PARAFAC1 or PARAFAC2) on vowel and subject components. The patterns for the

articulators indicate that there are strong differences between speakers with respect to the most influential measurement as revealed by PARAFAC2: In particular, the most influential y-contribution is the tongue-back for some talkers and the tongue-dorsum for other speakers. With respect to the speaker weights, again, the leverage patterns are very similar for both PARAFAC-versions. These patterns converge with the results of the loading plots, where the articulator profiles seem to be most altered by the use of PARAFAC2. These findings, in general, are interpreted as evidence for the reliability of the PARAFAC1 speaker weights.

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. Nevertheless, the exact purpose of their application is not as homogeneous as it might seem at first glance: It has been suggested to evaluate statistical articulatory models in terms of their potential to mimic articulatory behavior expressed in terms of articulatory degrees of freedom as in the tradition of Maeda (1979a,1979b,1990). But also, a second tradition has focused its attention on these methods' ability to generalize over several speakers . Likewise, the statistical procedures are tuned to different rationales: The first tradition leads to intraindividually fitted models advantageous for control purposes, as applied in articulatory control models for speech synthesis (Badin et al., 2002) or facial animation (Maeda, 2005). Multispeaker solutions are characterized by the attempt to reveal latent building blocks underlying articulatory organization. In this work, we will concentrate on the latter approach, i.e. the "PARAFAC-tradition".

1.1 Classical PARAFAC1

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)

$$X_k = AS_kV^T \tag{1}$$

where X_k is the kth "slab" of the input data matrix, with k the number of speakers, A is the matrix of articulator loadings, S is the diagonalized matrix of speaker loadings for speaker k and V the loading 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, citing 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." Fitting the PARAFAC1 to the data in the least squares sense amounts to minimizing

$$\sigma_1(A, V, S_1, \ldots, S_k) = \sum_{k=1}^{k} ||AS_kV^T||^2 \tag{2}$$

There is a unique solution minimizing (2) up to scaling and permutation. Cattell's proportionality 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.

Before turning to the constraints that define PARAFAC1 and describing less restrictive alternatives, we give a brief review of the studies using this method.

1.2 Survey of studies using PARAFAC1

The presumably largest focus of interest in the late 80's to the mid 90's by researchers from the speech production area using multimode Data Analysis techniques has been an issue raised in a paper by Jackson (1988) concerning the number of factors that are reliably extractable by means of the PARAFAC1 algorithm. Jackson claimed to have extracted a three-factor solution from a

corpus of Icelandic data. This claim was rejected later by Nix et al. (1996) highlighting the importance of diagnostic measures for the assessment of reliability of a PARAFAC solution: Harshman & Lundy (1984) suggested to use the triple product over the three modes of the correlations between corresponding sets of weights for each pair of factors. This triple product, also referred to as "congruence" coefficient can, in the case of PARAFAC1, be calculated as the triple Hadamard product of the products of the component matrices with their transposes:

$$TC = (A^T A) \circ (B^T B) \circ (C^T C) \tag{3}$$

Harshman & Lundy (1984) suggested triple products more negative than -0.3 between a pair of factors indicating a degenerate solution since in this case both factors are attempting to capture similar portions of the total variance, resulting in a second factor being simply a degenerate version of the first. The reanalysis of the data published by Jackson (1988) carried out by Nix et al. (1996) indicated that the third factor in Jackson's solution was not reliable which lead to disenchantment about the explicative claim made by this kind of modeling.

A second major result of the discussion of the 80's and 90's was the format of input data for applications of the PARAFAC1 algorithm to articulatory problems: "Although measuring the shape of the tongue with respect to anatomically normalized vocal diameter gridlines[1] does reduce the initial representational dimension, this measurement scheme needlessly loses 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" (Nix et al., 1996, p. 3708). In other words, the quality of the data seen by the algorithm determines the solution obtained by fitting PARAFAC, and therefore also the interpretation of the factors. This in particular can become a hot topic concerning the relevance of analyzes obtained by this method considering the advent of three-dimensional acquisition techniques in speech production research.

The first application of the PARAFAC algorithm in a reviewed journal contribution to three-dimensional tongue configurations was published in a paper by Zheng et al. (2003). The essential novelties apart from the three spatial dimensions of the input data consist in (a) a more thorough discussion of rea-

[1]The original PARAFAC work was based on the measurement of distances along anatomically defined reference lines forming a "measurement grid", which was calculated for each speaker

sonable preprocessing strategies for the application of the algorithm to tongue configurations and (b) the assessment of the solutions obtained by PARAFAC1 by more recent diagnostics of model degeneracy. With respect to the first point, the authors apply additional scaling subsequent to centering as applied in previous studies. The purpose of the scaling procedure is to normalize each speakers's data to unit sum-squared variation, "so that talkers with greater variability and/or larger vocal tracts do not dominate the PARAFAC fitting process" (Zheng et al., 2003, p. 482).

With respect to the second point in the preceding paragraph, i.e. the application of more recent diagnostics of the reliability of model fits, it is useful to have a closer look at family relationships between N-way methods. Here, PARAFAC1 can be considered as a special case of a more general method of three-way factor analysis, Tucker3 (Tucker, 1966). The structural model of Tucker3 is given in formula (4):

$$\underline{X} = VG(S \otimes A)^T. \tag{4}$$

Here, $\underline{X}$ denotes the higher-way array to be modeled, V, S and A are the component matrices (S the speaker weights, A the articulator weights and V the vowel weights). G denotes the so-called "Tucker core" matrix. $|\otimes|$ denotes the Kronecker tensor product. Now, PARAFAC1's structural model implies a hypercube as shaping of the core array , e.g. for a 2-factor solution the core array has the dimension 2 x 2 x 2. Furthermore, all elements off the hyperdiagonal of the core array are required to be zero for a valid PARAFAC solution, i.e. the core array is required to exhibit superidentity and therefore cancels in the following representation of PARAFAC1:

$$\underline{X} = V(S| \otimes |A)^T. \tag{5}$$

Here, $S| \otimes |A$ denotes the Khatri-Rao product. This conceptualization of PARAFAC is used in Bro (1998) for the development of an alternative criterion of the number of factors and the detection of model degeneracies in PARAFAC1 models. It measures the percentage of the variation in the Tucker core matrix G consistent with PARAFAC1's requirement of core hyperdiagonality. Bro & Kiers (2003) suggest that a core consistency of at least 90% is a good indicator of a valid model.

1.3 PARAFAC2

Above, we have mentioned Cattell's notion of "parallel proportional profiles", which does not always have to be a valid assumption; it can also turn out to be too restrictive in some cases, and, as we have shown elsewhere (Geng & Mooshammer, 2000), a less restricted algorithm, PARAFAC2, offers an attractive alternative. Referring to the notion we have used in equation (1) for PARAFAC1 for a single "slab" of the multiclassified array, PARAFAC2 can be expressed as

$$X_k = A_k S_k V^T \tag{6}$$

Within PARAFAC2, each loading matrix for the articulators, A_k, is expressed as $A_k = P_k A$. P_k is an $I * R$ matrix, where R denotes the number of factors and I the number of measurements in the articulator domain. A is constant over all these individual profiles and of size $R * R$. The rotational freedom provided by the PARAFAC2 model is adequate for approximating certain deviations from the strict linearity required in PARAFAC1. PARAFAC2 incorporates an invariance constraint on the factor scores as a milder version of factorial invariance: The cross-product matrix $A_k^T A_k$ is constrained to be constant over k speakers. The model structure is determined by the choice of the structure of Ak. Bro (1998) compares PARAFAC2's flexibility in this respect to Procrustes analysis. In Geng & Mooshammer (2000) we have shown that the strict assumptions required in the classical PARAFAC1 model were too strong to capture stress-specific variation in full detail. In contrast, PARAFAC2 allowed to account for systematic variation produced by word stress by imposing this weaker structure on the data. In particular, PARAFAC2 modeled the physical properties of the vocal tract shape in a more realistic and plausible way with respect to the description of mean factor shapes.

2 Method

2.1 The Corpora

In this study, we will reanalyze two distinct corpora. Both of them sample vowel nuclei acquired with fleshpoint tracking methods. The first corpus, which we will term the "stress corpus" was described in Geng & Mooshammer (2000), the second corpus, which we will refer to as the "speech rate corpus" was published in a paper by Hoole (1999). We will reiterate the description of these in order to pinpoint the differences between them, which could potentially endanger our interpretations concerning the method comparison we wish to achieve.

2.1.1 The Stress Corpus

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,a,ɑ,o,ɔ,u,ʊ/) 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 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 head movements under the helmet during the recording session. Jaw and lower lip movements will not be included in the analysis.

2.1.2 The Speech Rate Corpus

This corpus consists of seven adults, six males and one female. The experimental conditions were similar with respect to apparatus, tongue sensor placement, vowel environment and preprocessing to the stress corpus described in the previous section. The test utterances were formed by inserting the vowels into three different consonant contexts /p_p/,/t_t/ and /k_k/. Each symmetric CVC sequence was embedded in a carrier phrase with the structure *Ich habe geCVC gsagt (I said _).* The subjects were tested in two separate recording sessions, usually a few days apart, which lasted about one hour each. In the first recording session the speakers produced the utterances at normal speech rate, in the second recording at a fast speech rate.

2.1.3 Potential Problems

- The data of the speech rate corpus, in contrast to the stress corpus, were recorded on two different occasions. Therefore, the sensors had to be attached twice, potentially resulting in artifacts of sensor placement.
- The material of the stress corpus contained two test words per item. It does not seem implausible to assume that the amplitudes of articulatory movements reduce over the course of the intonation phrase.

3 Results

For both corpora, we performed two analyzes of the data, the first using PARAFAC1 and the second PARAFAC2. The analysis of the rate corpus concerning PARAFAC1 is a partial reanalysis of results published in Hoole (1999) and therein referred to as the model for "multiple consonantal contexts". Therefore, reprinting the displays already published in the paper mentioned would be redundant and is skipped with reference to the original publication. To stay in line with the results published in this paper, we also used the same preprocessing strategy as in Hoole (1999): The data delivered to the algorithm consisted of displacements from the average articulatory configuration of each subject. This amounts, in "standard terminology" (Harshman & Lundy (1984)) to centering across the vowel mode. This does not necessarily have to be the optimal preprocessing strategy, as elaborated in Zheng et al. (2003)[2], but was adopted here for optimal comparability. The same strategy was applied to the stress corpus as well, for comparability purposes. Note that beforehand, in Geng & Mooshammer (2000), we had applied centering across vowel and speaker mode[3], so that the solutions are not directly comparable to these results. Furthermore, as mentioned in Geng & Mooshammer (2000), we had to constrain some modes in some models to orthogonality in order to obtain non-degenerate solutions.
The results section is organized as follows: In the first part, we will have a look at global fit measures like the percentage of variance explained in order to achieve some basic insight into the structure of the models and to substantiate our solutions as valid. In the second, descriptive part, we display the conventional results on the solutions obtained, i.e. loading plots of extracted factors. In the third part, we proceed with analyzing the leverages, i.e. the influences that determine the exact solutions and the differences in fit between them.

3.1 Global fit

In the first step, we will have a look at the global measures for the different solutions. As mentioned above, some of the models were constrained to orthogonality in the vowel mode in order to prevent strongly correlated factors and degeneracy. This holds for both solutions analyzing the stress corpus, and

[2]We crosschecked the congruence between differently preprocessed factor solutions, more precisely between the strategy adopted here and the strategy recommended by Zheng et al. (2003) with additional scaling in the speaker mode. This measure resulted to 0.99 and evidences an almost identical solution.

[3]contrary to the citation in Zheng et al. (2003).

for the PARAFAC2 model of the rate corpus[4].
Note that, unlike in principal component analysis, the sum of the variances explained by single factors does not necessarily have to sum up to the total percentage of variance explained by the whole model. Table 1 summarizes these statistics. The percentage of variance explained for the PARAFAC1 in the speech rate corpus was around 80% , as already published in Hoole (1999). The amounts explained for the first and second factors amount to 61% and 24% . The fit of PARAFAC2 with respect to this dataset is slightly better. For the whole model this amounts to 82% and for the single factors to 21% and 61% respectively.

Concerning the stress corpus, we observed 86% variation explained for the total PARAFAC1 solution and 69% and 17% for the two factors separately. For PARAFAC2, the same indicators amount to 90% ,18% and 72% . Taken together, for this corpus, the benefit in explained variances by using PARAFAC2 was substantial in contrast to the speech rate corpus. The core consistency diagnostic can only be calculated for PARAFAC1 model and can take values less than or equal to 100. According to (Bro & Kiers, 2003, p. 276), a core consistency close to 100% implies an appropriate model, and, as a rule of thumb, a core consistency above 90% can be interpreted as 'very trilinear'. Accordingly, the consistency for both solutions reported here can be seen to almost perfectly conform to the PARAFAC1 model. In other words, valid solutions seem to be warranted and we can turn to the display of conventional loading plots.

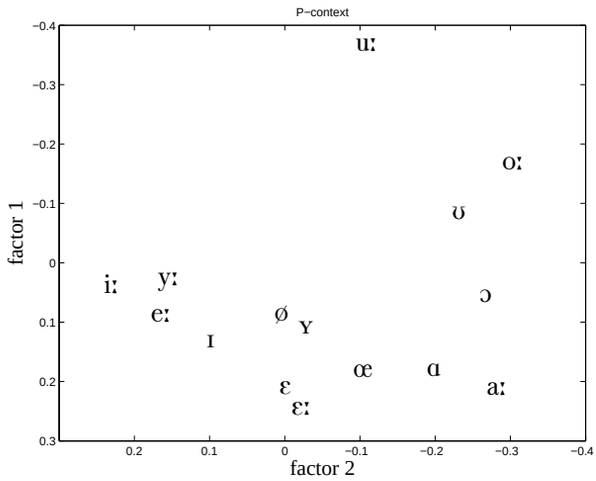
Table 1. Summary statistics for fitted models

	Perc.expl.tot	**Perc.expl.F1**	**Perc.expl.F2**	**CoreCond**	**Congr. tot**
P1 rate	80%	61%	24%	100	-0.05058
P2 rate	82%	21%	61%	-	1
P1stress	86%	69%	17%	98.3	0.00008
P2 stress	90%	18%	72%	-	1

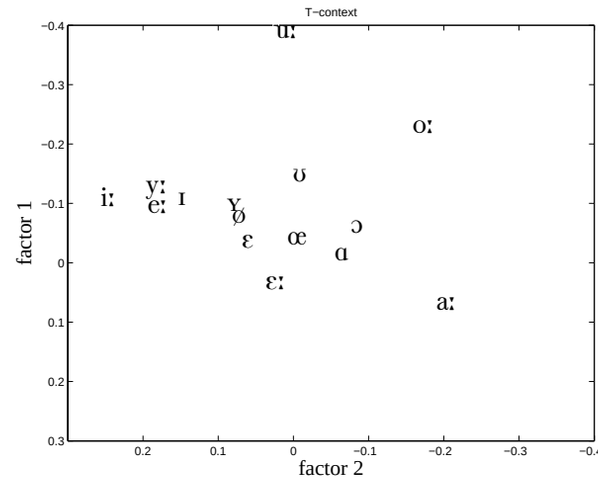
3.2 Loading Plots

As can already been seen from table 1, the ordering of the factors is reversed in the PARAFAC2 solutions resulting in a second factor with a higher percentage of variance explained than the first factor. For the plots of vowel loadings, the axes were changed according to convention, i.e. with high front vowels in

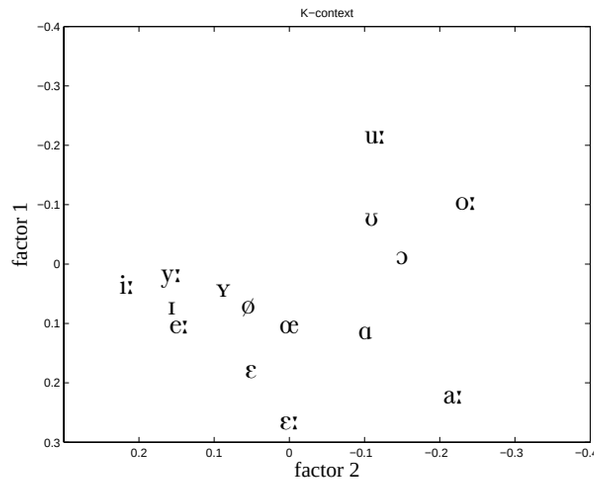
[4]Note that constraining the first (vowel) mode to orthogonality implies constraining the second (articulator) mode. Nevertheless, congruences of around .95 for the unconstrained speaker modes were indicating non-degenerate models



(a) PARAFAC2: rate database, P-context



(b) PARAFAC2: rate database, T-context



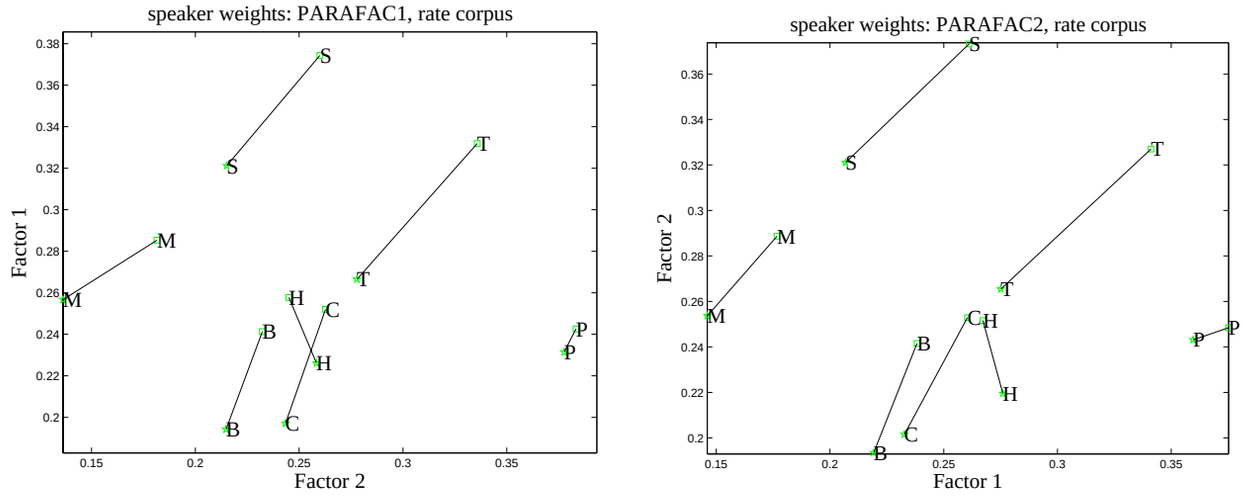
(c) PARAFAC2: rate database, K-context

Figure 1. Results for the *SPEECH RATE* database for whole corpus, vowel loads

the top left corner. This is not in general the case for the plots of the speaker weights.

3.2.1 Speech Rate Corpus

Figure 1 shows the loading plots for the vowel mode split by consonantal contexts for the speech rate corpus. These plots can be directly compared to figure 4 in Hoole (1999). The PARAFAC2 solution can be seen as a rotated version of the vowel space as described in Hoole (1999), i.e. the topological information is retained. This implies that we do not have to discuss this aspect in further detail. Similarly for the speaker weights shown in figure 2. Here, the results for the PARAFAC1 solution are identical with the results of figure 6 in Hoole (1999). Both solutions conform to a scaling down of articulation for subjects B, C, M, S, and T in the fast rate condition, and a different behavior for speakers H and P, conforming to the fact that an increase in speech rate can be achieved by either downscaling the amplitudes of articulatory movements or by increasing movement velocities. For the current study the absence of a substantial and interpretable topological change between the two solutions is the interesting aspect.



(a) speaker space rate, PARAFAC1 (b) speaker space rate, PARAFAC2

Figure 2. Results for the *SPEECH RATE* database, speaker weigths

In contrast to Hoole (1999), we show the loading plots for the articulator weights split by paralinguistic conditions. These plots can be seen as the effects of the factors on the tongue configurations of an average speaker. For the speech rate corpus, this information is given in figure 3. In general, both solutions cohere with the interpretation of the factors published in Hoole (1999). The

most striking result in the PARAFAC1 solution appears to be the absence of a strong difference in tongue shape between the projections at normal and fast rate, the projection at fast rate being a somewhat downscaled version of the projection at normal rate. This is slightly different for the PARAFAC2

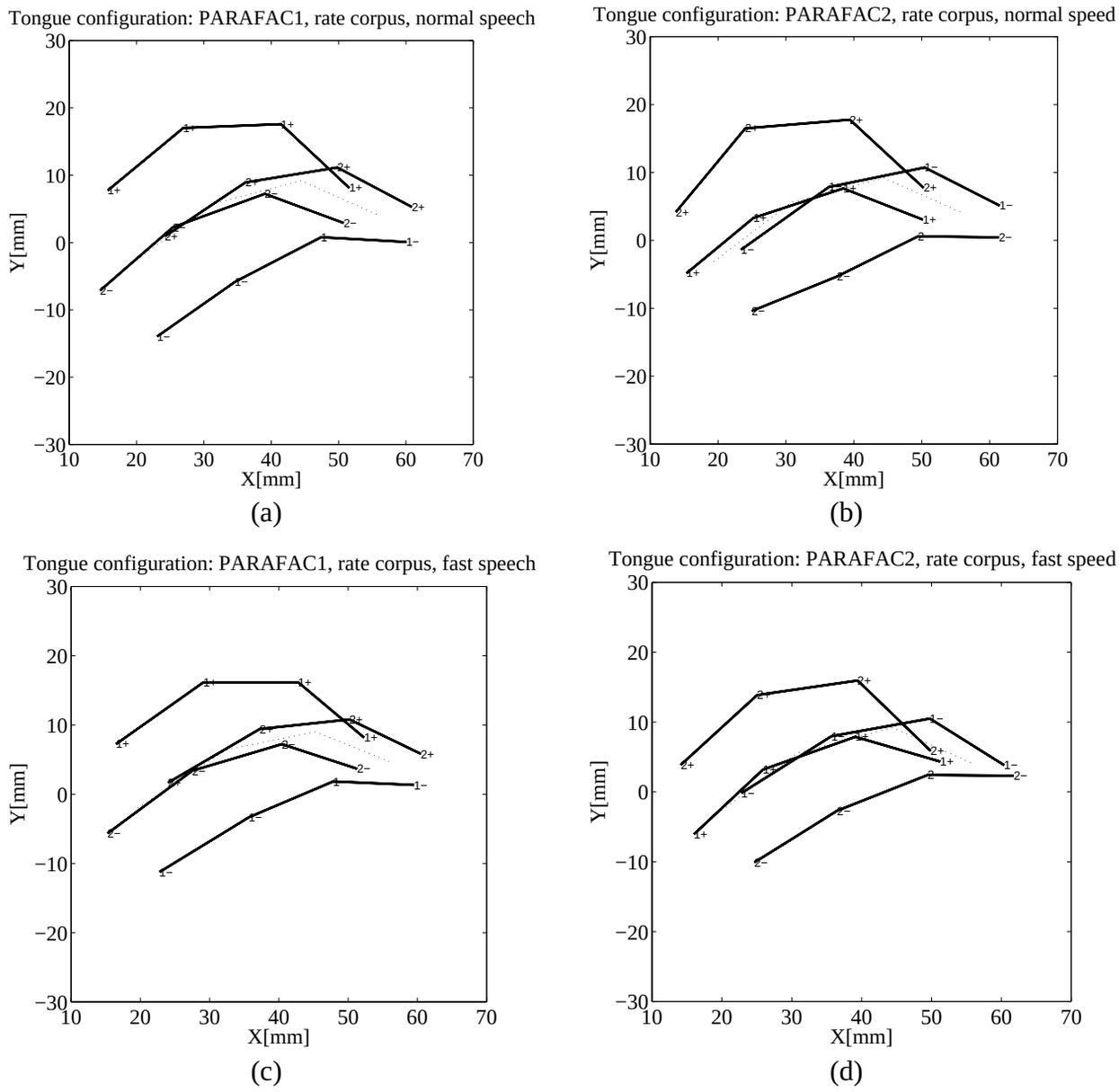
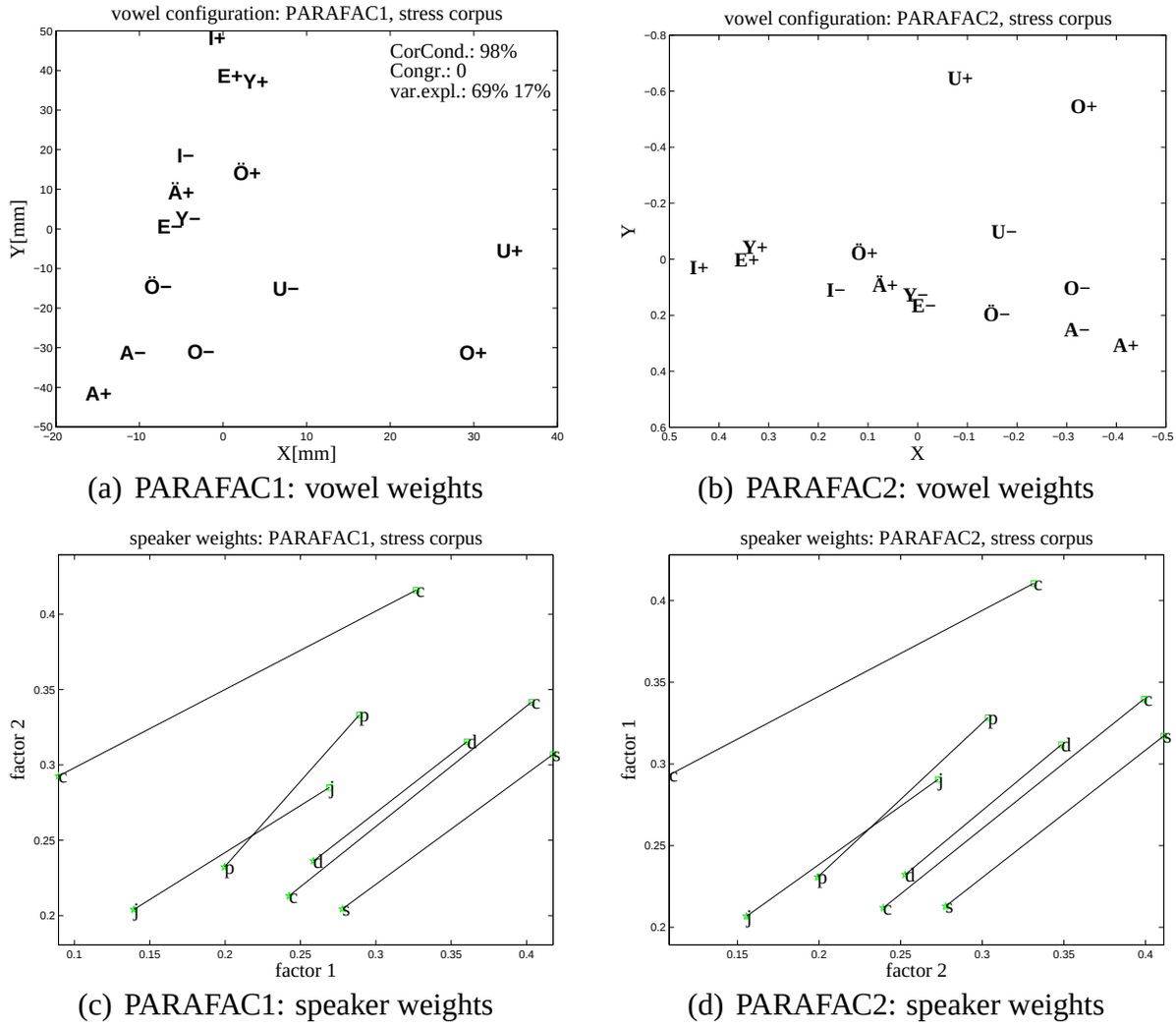


Figure 3. *speech rate* database, articulator loadings

solution. A downscaling of amplitudes is indeed observed, but additionally, there are some shape-relevant aspects in this solution worth mentioning: First, the negative shape of factor2 -corresponding to factor 1 in PARAFAC1, front-raising- in the normal-rate condition is characterized by a lower tongue blade sensor in comparison to the surrounding tongue tip and tongue dorsum sensors. If this factor is assumed to encode a movement from an /a/-like to an

/i/-like shape, the /a/-pole of this factor appears to be a more reasonable configuration than the first factor of the PARAFAC1 solution. Subsidiary evidence comes from the comparison of the /i/-like pole of the front-raising factor at normal speech rate: The tongue tip appears to be "more down" in the PARAFAC2 solution, which as well seem to be more reasonable. Interestingly, the informative patterns with respect to factor 2 arise in the fast-rate condition. Hoole (1999, p.1026) had noted that his second factor shares with the solution found in Harshman et al. (1977) " the responsibility for forming a constriction in the velar region, but our factor 2 shows above all a pattern of advancement and retraction, which is hardly the case for the "back raising" factor." This tendency



(a) PARAFAC1: vowel weights

(b) PARAFAC2: vowel weights

(c) PARAFAC1: speaker weights

(d) PARAFAC2: speaker weights

Figure 4. Results for the Stress database. Left Panels: PARAFAC1, right panel: PARAFAC2. Tense vowels(+), lax vowels(-)

appears to be even more prominent for the tongue back sensor in the fast rate condition for the PARAFAC2 solution, where no raising movement at all is ob-

servable.

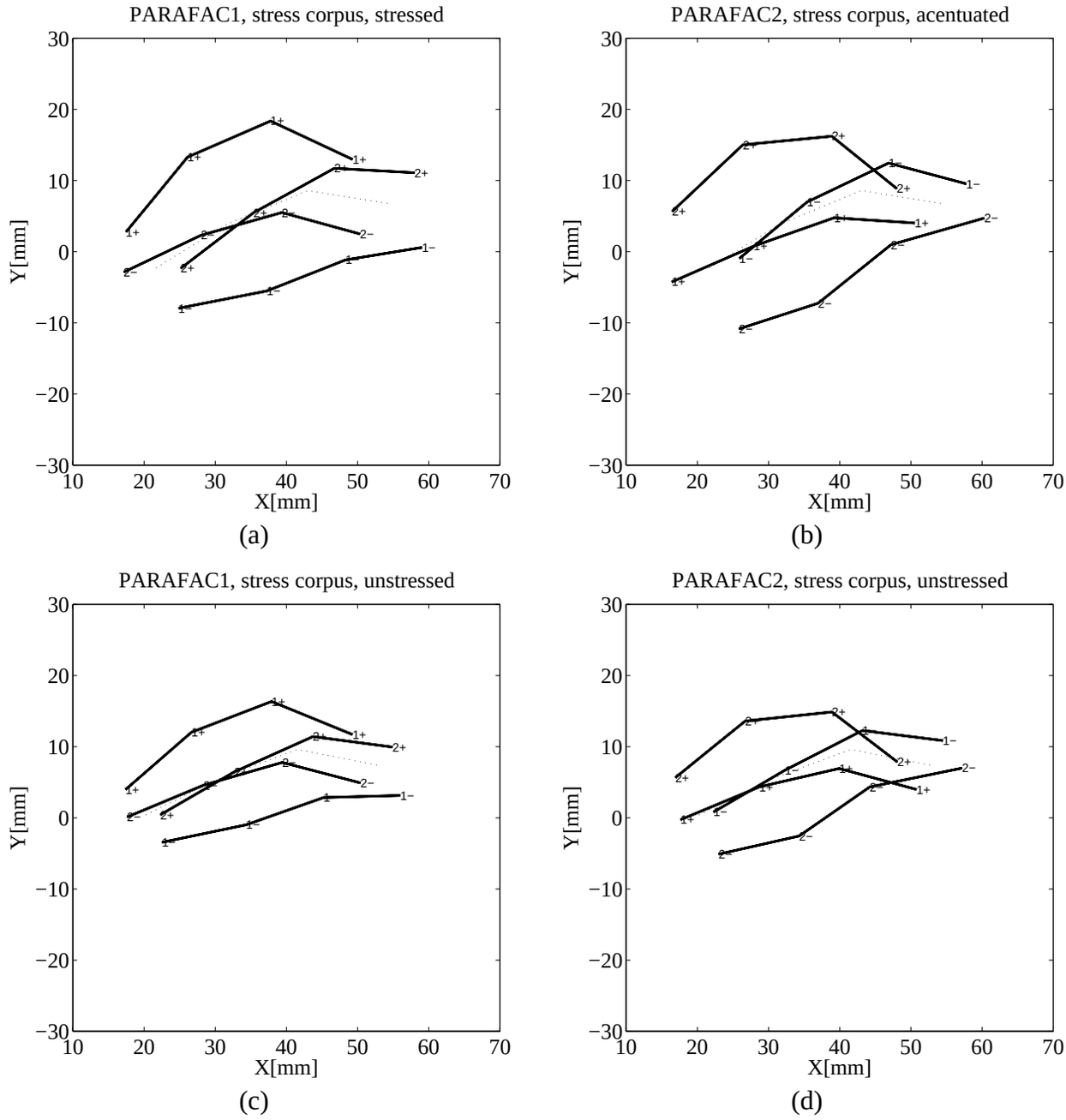


Figure 5. Articulatory configurations for the stress database

3.2.2 Stress Corpus

As mentioned above, for the stress corpus, both models were constrained to orthogonality in the vowel mode. The PARAFAC1 model was quite close to degeneracy, but we have shown the core consistency above (table 1) indicating an acceptable coherence with trilinearity. In short, a pattern comparable to the one

for the speech rate corpus was observed concerning vowel and speaker weights shown in figure 4: There is no evident change in topology in vowel and speaker plots comparing PARAFAC1 and PARAFAC2. With respect to the articulatory configurations, the patterns are partly similar to the speech rate corpus: The "trough-like" -shape of the tongue blade is also evident for /a/-like configuration, but is visible in both stress conditions. Interestingly, both PARAFAC2 factors have quite a strong horizontal component except for the back raising factor in unstressed condition.

3.2.3 Preliminary Summary

In this paragraph, the results obtained until so far will be summarized. PARAFAC1 and PARAFAC2 solutions give comparable results with respect to speaker weights and vowel weights. This kind of topological invariance could be substantiated in a more formal way by showing that e.g. the shape difference in the vowel spaces of PARAFAC1 and PARAFAC2 solutions is uniform, i.e. only trivial translation, scaling, and rotation operations are involved. This idea is not tracked further here, but could be performed by Generalized Procrustes analysis (Gower, 1975). The articulatory configurations for the "modal speaker" with respect to the paralinguistic features seem to show enhanced "flexibility" for PARAFAC2. Nevertheless, the gain in variance explained is only substantial in the stress corpus - the "/t/-only" data set. In the next paragraph, we will apply a method to identify the most influential observations shaping the particular solutions, particularly with regard to possible biases in the speaker weights caused by the compromise quality of the PARAFAC1 solution.

3.3 Leverages

Leverages were originally developed for regression analysis as a tool for residual and influence analysis. For this reason, it might be more appropriate to speak of the "squared Mahalanobis distance" in the context of a factoring method like PARAFAC. Anyway, leverages are also widespread in two-way Principal Component Analysis, therefore the term "leverage" also seems appropriate (Bro, 1997). For a particular loading matrix, e.g. for first mode loadings, leverages can be calculated as

$$v = diag(A(A'A)^{-1}A') \tag{7}$$

Their possible range is between 0 and 1. A high value indicates that an observation is influential, a low value indicates the opposite. As mentioned already, leverages could have been calculated for each of the two factors separately for

every mode. Here, we will limit ourselves to their evaluation with respect to the factor structure as a whole. The basic results with respect to the leverages in the vowel mode point to a corpus effect: The least influential observations are the palatal vowels, the strongest contributions are made by long back vowels and /a:/ This presumably is a corpus effect of the structure of the German vowel system with its numerous front vowels. Furthermore, lax vowels are generally less influential than their tense counterparts. The leverages in the speaker-mode show hardly any effect of the paralinguistic variables: Fast rate and unstressed shapes are generally less relevant for the total solution than normal rate and stressed shapes. This holds with the exception of speaker H, who is characterized by a deviant articulatory implementation of speech rate Hoole (1999).

The calculation of leverages in the articulator mode offers an additional inter-

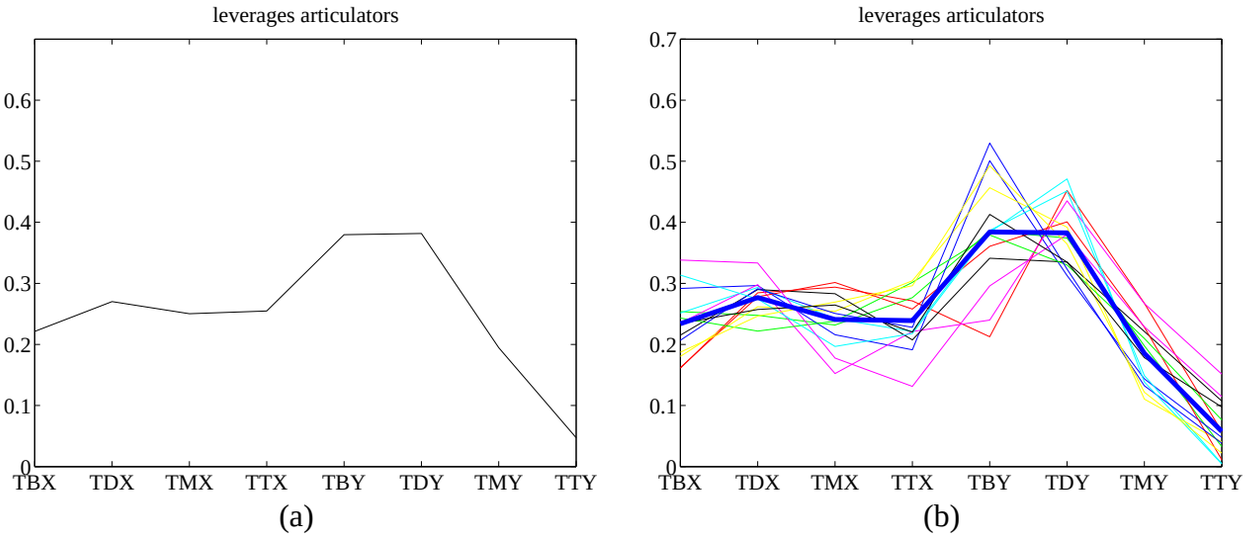


Figure 6. rate corpus: Leverages articulator mode, Sensor naming conventions: TB:tongue back, TD: tongue dorsum, TM: tongue mid, TT: tongue tip. x and y denote horizontal and vertical components.

esting property: Whereas for PARAFAC1, only one leverage profile can be calculated for the matrix of articulator weights, PARAFAC2 offers the possibility of calculating leverages for each speaker separately. Figure 6 and 7 show these plots for the two data sets used in this study. The left panels show the leverages for the articulator weigths of the respective PARAFAC1 solution, the right panels show the leverages for the individual speakers as obtained by PARAFAC2. The bold line in the right panels depict the average values of the single speaker's articulators for PARAFAC2.

The most evident patterns of figures 6 and 7 are (a) the general peak in influence observed for tongue back and tongue dorsum in y-direction (b) the su-

perficially close similarity between the average leverage profile of PARAFAC2 and the corresponding PARAFAC1 profile. This correspondence is almost perfect in the rate data set. Contrastingly, for the stress data set, a shift of the most important sensor for tongue dorsum to tongue back can be observed for PARAFAC2 in comparison to the PARAFAC1 profile.

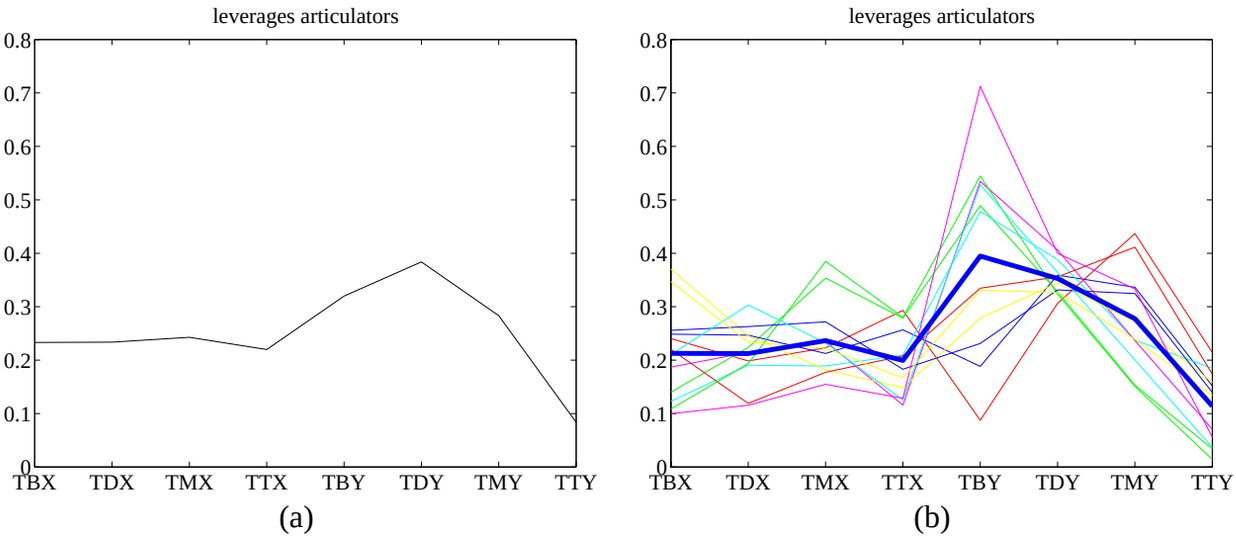


Figure 7. Stress corpus: Leverages for articulators. For explanation of the x-axis see figure 6.

Taken together, the patterning of the different modes in the leverage analyses is equivalent to the patterning in the loading analyses in previous sections with respect to the factoring method. This raises the following question: If the greater flexibility of PARAFAC2 to account for interindividual differences in the articulatory configurations by a separate subspace for each speaker does not relevantly influence speaker and vowel spaces, what is the origin of the increase in fit observed for the stress data set? In order to answer this question, we calculated explained percentages of variance for each speaker separately across data sets and factoring method. The result is shown in figure 8. The results a general increase in explained variances for PARAFAC2 for both paralinguistic features in both data sets. Conforming with the results on the explained variances for the total samples, this increase in fit is less pronounced for the speech rate than for the stress corpus.

4 Discussion

The results of these analyses can be summarized as follows: When comparing two different PARAFAC-versions on two different corpora, findings with respect to the explained variances are better for the PARAFAC2 algorithm for

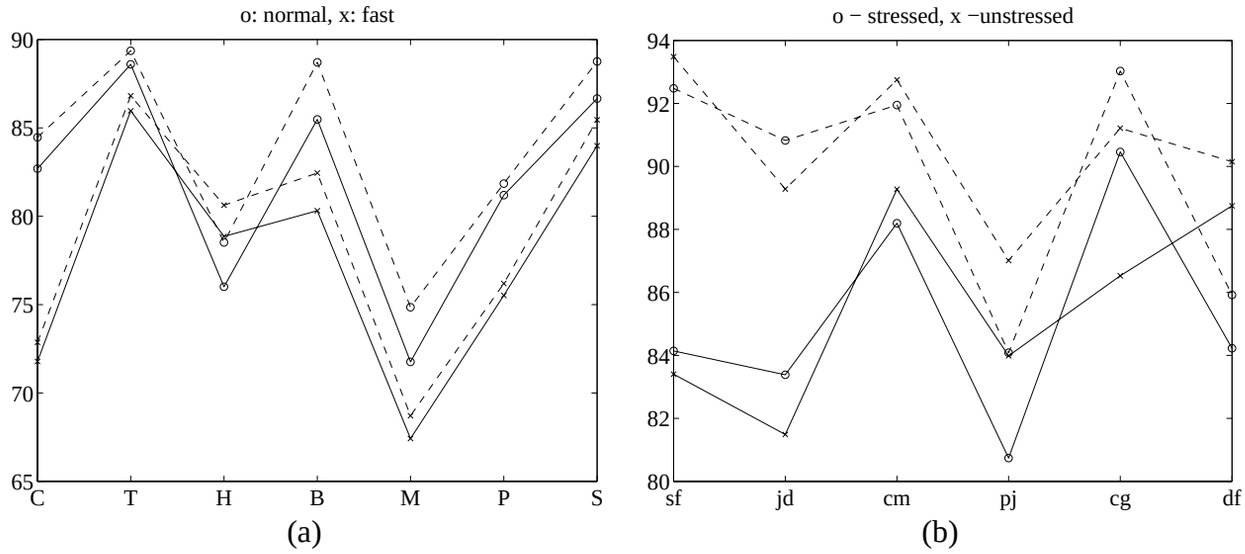


Figure 8. Percent explained per speaker for rate corpus (a) and for stress corpus (b). dashed lines. dashed lines: PARAFAC2, solid lines: PARAFAC1. left panel (a): circles indicate normal rate, crosses fast rate. right panel (b) circles indicate stressed, crosses unstressed position

both corpora. Note that the smaller impact of the choice of the modeling strategies on vowel and especially speaker weights is not precluded by the algorithmic decision made; rather it originates in the data analyzed. This is evidenced by the relatively large impact of interindividual differences on the leverages in the speaker modes in PARAFAC2 for both corpora.

We mentioned in section 2.1.3 that the data of the speech rate corpus were recorded on two different occasions implying that the sensors had to be attached twice. This does not appear to be problematic. The solution of the speech rate corpus is more stable than the solution of the stress data set: A scenario in which sensor placement differences between sessions would be recovered in individual articulator spaces would have been easily conceivable. However, this is not the case, as the speech rate data set benefits less from PARAFAC2 in terms of explained variances than does the word stress data set. Therefore, it seem to be conclusive that the greater impact of PARAFAC2 on the solution in the word stress data set can be traced to the coarticulatory influences of alveolar articulation present in these data. In this sense the word stress data set analysed here confirms the findings of Hoole (1999), where analyses of single consonant contexts ran into problems for the alveolar context. This is nicely illustrated by leverages of vowel mode loadings where the most influential observations stem from long back vowels. One need not go as far and claim that the canonical 2-

factor PARAFAC solution is largely incompatible with vowel data acquired in alveolar contexts, but often stabilizing orthogonality constraints have to be applied in order to end up with an interpretable model. In these cases, PARAFAC2 can capture speaker-specific variation more accurately than PARAFAC1.

This converges with the finding that the tongue back / tongue-dorsum y-components are the most influential observations in the data sets. If the vowel shapes are influenced by a surrounding /t/-gesture, the tongue-dorsum and tongue-back becomes articulatorily more constrained during the vowel. This leads to a decrease in the variance generated in the "backward-upward" direction and an increase in variance in "front raising"-conform direction. In other words, the "dominance" of the "front-raising" factor increases leading to a fragile second factor or even degeneracies. We think that the Procrustes-like relaxed version of the Parallel Proportional Profile as implemented in the PARAFAC2 algorithm allows for a speaker-specific definition of this backward-upward movement and prevents degeneracy as shown in the stress data set. There are other situations where this "backward-upward direction" of the tongue variance could be ill-defined across speakers and where PARAFAC2 could be a remedy against methodological pitfalls: Using "fleshpoint methods" like the magnetometer, the experimenter decides for particular landmark definitions while gluing the sensors on the tongue. Acquiring data using different methods like the three-dimensional reconstruction of MRI data as described in the paper by Zheng et al. (2003), the input data are the outputs of surface reconstruction algorithms, and the definition of landmarks is carried out at a later stage in the analysis. Here, the analysis could benefit from the fact that PARAFAC2 fits the data directly and not cross-products between column units. Therefore, it is easier to handle missing data, and in the case of three-dimensional data, the strong concept of the landmark is to some extent relaxed. It would be possible to let the dimension of the articulator mode differ from slab to slab, hence each slab k could have its own specific articulator mode dimension and thus the solution potentially depends less on a particular landmark definition.

Zheng et al. (2003) seem to have encountered morhological problems and solved them by applying advanced preprocessing strategies consisting in an implicit vocal tract length normalization (see section 1.2). We crosschecked the preprocessing recommendations published in Zheng et al. (2003) against the word stress corpus applying the vowel centering we were using in this study and in Hoole (1999). This led to practically identical solutions compared to the PARAFAC1 solution we reported in this study. For three-dimensional data however, such a preprocessing approach might be of greater benefit than for the analysis of EMA corpora, because static three-dimensional data might em-

phasize the importance of vocal tract morphology. But at the same time, in static MRI settings, the speaker weights might be more strongly biased by these methodological problems. At the moment, we only can state, that for our EMA analyses reported here, speaker weights were remarkably stable, although we still cannot offer a conclusive interpretation on what they measure.

References

Badin, P., Bailly, G., Rveret, L., Baciu, M., Segebarth, C., & Savariaux, C. (2002). Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, 30:533–553.

Bro, R. (1997). PARAFAC. tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38:149–171.

Bro, R. (1998). *Multi-way Analysis in the Food Industry Models, Algorithms, and Applications*. Ph.D. thesis, Department of Dairy and Food Science Royal Veterinary and Agricultural University Denmark.

Bro, R. & Kiers, H. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17:274–286.

Geng, C. & Mooshammer, C. (2000). Modeling the german stress distinction using PARAFAC2. In: *Proc. 5th Speech Production Seminar*, 161–164.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.

Harshman, R. & Lundy, M. (1984). Data preprocessing and the extended PARAFAC model. In: H. Law (ed.) *Research Methods for Multimode Data Analysis*, 216–284. New York: Prager.

Harshman, R. A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62:693–707.

Hoole, P. (1999). On the lingual organization of the german vowel system. *Journal of the Acoustical Society of America*, 106:1020–1032.

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

Maeda, S. (1979a). An articulator model based on a statistical analysis. In: J. Wolf & D. Klatt (eds.) *Speech Communication Papers*, 67–70. Acoustical Society of America, NY.

Maeda, S. (1979b). Un model articulatoire de la langue avec des composantes lineaires. *10mes journes d'tudes sur la Parole, Groupe Communication Parle*, 152–164.

Maeda, S. (1990). *Speech Production and Speech Modeling*, chapter Evidence from the Analysis and Synthesis of Vocal Tract Shapes using an articulatory Model, 131–149. Behavioural and Social Sciences. Hardcastle, J. and Marchal, A.

Maeda, S. (2005). Face models based on a guided pca of motion-capture data: Speaker dependent variability in /s/-/?/ contrast production. *ZAS Papers in Linguistics*, 40:95–108.

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*, 99:3707–3718.

Tucker, L. (1966). Some mathematical notes on three-way factor analysis. *Psychometrika*, 31:279–311.

Zheng, Y., Hasegawa-Johnson, M., & Pizza, S. (2003). Analysis of the three-dimensional tongue shape using a three-index factor analysis model. *JASA*, 113:478–486.