On the lingual organization of the German vowel system

Philip Hoole
Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilians-Universität München, Schellingstrasse 3, D-80799 Munich, Germany

(Received 26 October 1998; revised 29 April 1999; accepted 3 May 1999)

A hybrid PARAFAC and principal-component model of tongue configuration in vowel production is presented, using a corpus of German vowels in multiple consonant contexts (fleshpoint data for seven speakers at two speech rates from electromagnetic articulography). The PARAFAC approach is attractive for explicitly separating speaker-independent and speaker-dependent effects within a parsimonious linear model. However, it proved impossible to derive a PARAFAC solution of the complete dataset (estimated to require three factors) due to complexities introduced by the consonant contexts. Accordingly, the final model was derived in two stages. First, a two-factor PARAFAC model was extracted. This succeeded; the result was treated as the basic vowel model. Second, the PARAFAC model error was subjected to a separate principal-component analysis for each subject. This revealed a further articulatory component mainly involving tongue-blade activity associated with the flanking consonants. However, the subject-specific details of the mapping from raw fleshpoint coordinates to this component were too complex to be consistent with the PARAFAC framework. The final model explained over 90% of the variance and gave a succinct and physiologically plausible articulatory representation of the German vowel space. © *1999 Acoustical Society of America.* [S0001-4966(99)03608-5]

PACS numbers: 43.70.Bk, 43.70.Kv, 43.70.Aj, 43.70.Jt [AL]

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 is 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 can certainly be regarded as involving a rich set of contrasts.

The search to understand the motor implementation of such a system can lead in a number of different directions. For example, there is the major question of the interarticulatory coordination of different speech organs. Thus a phonologically defined contrast such as rounding proves to involve not only labial activity but also positioning of tongue and larynx (Wood, 1986; Hoole and Kroos, 1998). In a similar vein, there is the question of how lingual and mandibular activity are coordinated for vowel articulation (Johnson *et al.*, 1993; Hoole and Kühnert, 1996). A second important direction concerns the temporal organization of speech, for example, the way in which a contrast such as tense versus lax is reflected in the organization of elementary CV and VC movements for the production of complete syllables (Kroos *et al.*, 1997). In this paper we will be concentrating on a third important area, namely on the search for an efficient and hopefully revealing characterization of *resulting* tongue position in vowel production (i.e., we leave aside the question of separate lingual and mandibular contributions to resulting tongue position). This is a further crucial level, since tongue shape is largely responsible for vocal tract shape and thus for fundamental acoustic properties of the sounds produced (see Hoole, 1999, for preliminary analysis of articulatory-acoustic relations based on some of the speech material used in the present study).

We will explore a data-driven procedure for deriving a model of vowel articulation. This approach seems justified given that no complete concensus exists for the most appropriate articulatory characterization of vowels (Wood, 1975; Fischer-Jørgensen, 1985). Nonetheless, 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.

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. Thus it is inherently very unlikely that each of the many German vowels represents a unique way of configuring the tongue; rather, one would suspect that vowels scale a few common underlying patterns in slightly different ways. Indeed, given the fact that many descriptions of vowels use a two-dimensional framework (e.g., the traditional vowel chart; classification in terms of location, and width of the main constriction; $F1$ versus various combinations of the higher formants) it would be fairly surprising if the number of dimensions determined from direct measurement of tongue shape were substantially different from two. But precise number and nature remain an open empirical question and cannot be assumed *a priori*.

We can also consider the question of empirically uncovering the organizational principles underlying observable articulatory behavior from the point of view of the raw data available to articulatory analysis. We will be working with fleshpoint data from EMMA sensors. The raw data from such a sensor are not particularly revealing in themselves; the simple act of gluing a sensor to the tongue, however, carefully and systematically done, introduces an element of arbitrariness to the data. But it is a common problem in psychological research, and one of the motivations for the development of factor analysis, that the underlying behavioral ‘‘building blocks’’ cannot be measured directly, but must be inferred from a multiplicity of (probably correlated) measures made on the behavioral surface. Similarly, direct measurements of the possible physiological building blocks of speech are very difficult to make, even with EMG (but see Maeda and 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 eight raw articulatory variables we have available in our data set (corresponding to two spatial dimensions measured at four sensor locations).

At the very least, such an endeavour should lead to a more readily apprehensible picture of the relation between vocal tract shape and linguistic structure, and ideally the results should be characterized by low dimensionality compared to the raw variables, phonetic interpretability, a potentially close relationship to the actual dimensions of organization employed by speakers, and finally by generalizability over speakers and perhaps languages.

We will here propose a hybrid PARAFAC and principal-component model of tongue position in German vowel production. The initial focus will be on the PARAFAC approach, which has given phonetically interesting results in a range of investigations (Harshman *et al.*, 1977; Jackson, 1988; most recently Nix *et al.*, 1996). PARAFAC is one of a class of three-mode analysis procedures, contrasting with standard principal-component and factor analyses, which are two-mode procedures. In the latter, the data to be analyzed are arranged in a two-dimensional array of observations (in our case the individual vowels) for a set of variables (in our case the fleshpoint coordinates). PARAFAC requires an inherently three-dimensional data structure, with the third dimension being represented in our case by the speakers (for a recent very extensive alternative approach to the analysis of multi-speaker datasets see Hashi *et al.*, 1998). The main 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. A further related advantage, which is particularly important in the context of our current main goal of understanding the articulatory structure of a complete vowel system, is that the linguistic identity of the utterances analyzed is directly reflected in the way the data are structured for input to the PARAFAC algorithm. In other words, the data structure implicitly captures the investigator’s knowledge as to what constitute linguistically equivalent observations for the different speakers. This contrasts with typical use of two-mode principal component or factor analyses (e.g., Maeda, 1990), where the aim is to sample the space of possible tongue shapes in some appropriate way, but without any particular reference to the linguistic identity of the selected observations.

Nonetheless, the PARAFAC model has a simple linear form:

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 three loading matrices $\mathbf{V}$, $\mathbf{A}$, and $\mathbf{S}$ (for vowels, articulators, and speakers) with dimensions $nv \times nf$, $na \times nf$, and $ns \times nf$, respectively.

For speaker k the complete dataset $\mathbf{Y}_k$ (dimension $na \times nv$) predicted by the model is then given by

$$\mathbf{Y}_k = \mathbf{A}\mathbf{S}_k\mathbf{V}^T, \tag{1}$$

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

The articulators could be either a set of measurements along predefined gridlines or a set of fleshpoint x and y coordinates. Measurements for these articulators are assumed to be expressed as deviations from the mean for each speaker over all vowels (the formulation given here follows Jackson, 1988, p. 129. See Nix *et al.*, 1996, p. 3708 and Harshman *et al.*, 1977, p. 699, for alternative notations).

The simplicity of the model should be apparent from this formulation. Its potential for a parsimonious representation can be illustrated as follows: If two factors are enough to model a hypothetical dataset of ten vowels, ten speakers, and ten articulators, then the total size of the loading matrices is $2\times(10+10+10)=60$ compared to $(10\times10\times10)=1000$ elements in the raw dataset.

Nonetheless, finding a solution to the PARAFAC equation is mathematically more complex than the two-mode case and experimenter judgement plays a greater role.

In particular, the algorithm must be told in advance how many components to extract, whereas for principal-component analysis one can simply decide afterward how many components to retain for further consideration. In addition, the reliability of the solution must be assessed: Jackson (1988) discusses criteria for successful solutions under the headings convergence, uniqueness, degeneracy, generalizability, and goodness of fit (each of these criteria will be expanded on where appropriate below).

Moreover, there are also 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 in the model; 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. As Harshman *et al.* (1977) put it:

‘‘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’’ (p. 699).

Are these assumptions justified for human speech behavior? Interestingly, more recent work from UCLA (Johnson *et al.*, 1993) seems to have seen a turning away from this model and an emphasis on speaker-specific articulatory strategies that would not be compatible with the PARAFAC model. This kind of behavior emerged particularly from an analysis of patterns of *inter*articulator coordination (tongue, jaw) but the authors concede that the assumptions of the model may still hold for an examination of *resulting* tongue position. Nix *et al.* (1996) appear to concur with this view:

‘‘the current claim is not that all speakers articulate the same vowels in exactly the same way; the claim made here is that two specific dimensions form an effective basis for the space of tongue shapes...’’ (p. 3716).

Thus while one may even go as far as Nix *et al.* (1996) that the model is ‘‘undoubtedly ultimately incorrect’’ (p. 3708), it has nevertheless consistently given phonetically interesting characterizations of vowel systems. Moreover, by applying this attractively simple model we obtain the important benefit of a *quantitative* estimate of what might remain to be gained—at the price of greater complexity—from a more sophisticated model, and thus of how urgent the search for such a model really is. In addition to this quantitative benefit, it indeed turned out in the course of applying the model that we obtained improved *qualitative* insight as to where, in phonetic terms, speaker normalization by the simple PARAFAC linear scaling approach is too restrictive. Specifically, this mainly appeared to involve consonantal influences on vowel articulation, and led to the abovementioned hybrid modeling approach, in which the PARAFAC model was supplemented by a principal-component approach that retained as much as possible of the spirit of the PARAFAC approach, while incorporating a relaxation of the constraints on the possible form of speaker-specific effects.

I. SCOPE OF THE INVESTIGATION

The dataset to which we wished to apply the PARAFAC approach is richer in two main respects than those reported elsewhere in the literature. First, we had recorded data for seven speakers. This is a larger number than has previously been used for PARAFAC analyses of tongue configuration (though Linker, 1982, has analyzed lip configuration for eight speakers of multiple languages). More significantly, each speaker recorded the speech material at two different speech rates (normal and fast) in separate sessions. It has been clear since the investigation of Kuehn and Moll (1976) that speakers implement an increase in speech rate by different means. The main possibilities appear to be either a general scaling down of articulation, or a pattern in which there is little target undershoot, but in which temporal compression (not considered directly here; see Kroos *et al.*, 1997) is achieved by increasing velocity. Both these patterns represent consistent types of articulatory behavior, which should emerge as such in the speaker weights derived by the PARAFAC algorithm. For the present purposes the main interest is methodological: The claim that the PARAFAC algorithm allows us to capture underlying principles of articulatory organization would be seriously compromised if speaker weights varied haphazardly over sessions. This would suggest that the algorithm is unduly sensitive to incidental but unavoidable differences in recording conditions over sessions. In practice, for the purpose of running the algorithm the seven speakers×two sessions are simply treated as 14 different speakers. After running the algorithm the patterns in the speaker weights for the two sessions can then be compared.

Second, our speech material also included all vowels in three different consonantal contexts. Previous PARAFAC analyses have typically analyzed vowels in only one context (or at least only one token per vowel). Would it be possible to capture effects of consonantal context on vowel articulation in this kind of analysis? As we will see below, this task turned out to be not completely straightforward and required a departure from the basic PARAFAC model.

Application to data with carefully controlled consonantal contexts was also a necessary first step toward potentially being in a position to apply the PARAFAC approach to a further more natural corpus we have available for each speaker, in which each vowel is spoken in 15 different consonantal contexts.

A further difference between our work and earlier work lies in the use of fleshpoint data. Nix *et al.* (1996) suggested that the original PARAFAC work based on cineradiographic measurements made along anatomically defined grid lines may artificially constrain the possible solutions—i.e., there is no straightforward way of capturing horizontal movement of tongue tip/blade. In their reanalysis of Harshman *et al.*’s (1977) radiographic data (13 gridlines) they determined the x/y coordinates of 13 ‘‘pseudo-pellets’’ equally spaced on the tongue contour (*op. cit.*, p. 3711) and suggested that the resulting solution was more easily interpretable. Yet as far as we know, directly measured as opposed to reconstructed fleshpoint data has not yet been analyzed with the PARAFAC technique, so the above claim could clearly benefit from further substantiation. Measured fleshpoint data also have one clear disadvantage compared to radiographic data, which is that the pharyngeal region is typically not well represented. However, work by Kaburagi and Honda (1994) using simultaneous articulographic and ultrasound measurements of the tongue indicated that the tongue contour could be reconstructed quite well from electromagnetic sensors attached to the tongue at realistically accessible locations (see also Badin *et al.*, 1997).

In purely numerical terms we have 8 pieces of articulatory information available per utterance (4 sensors×2 coordinates), compared to 13 for the original Harshman *et al.* radiographic study.

A final minor point where our work supplements previous work is that the German vowel system has yet to be analyzed with this approach. The German vowel system differs both phonetically and phonologically quite substantially from the American English system. In particular, due to the presence of front rounded vowels (and leaving diphthongs out of consideration) there are 50% more vowels to be considered.

II. THE DATASET

A. Subjects

The speakers consisted of seven adults, six males and one female, all phonetically trained and experimentally experienced. Their dialects showed no marked regional characteristics but conformed to general High German.

B. Speech material

The speech sample consisted of all (15) monophthongal vowels of German. These can be grouped with one exception (ɛː) into tense–lax (long–short) pairs: /iː,ɪ/, /yː,ʏ/, /eː,ɛ/, /øː,œ/, /ɑː,a/, /oː,ɔ/, /uː,ʊ/.

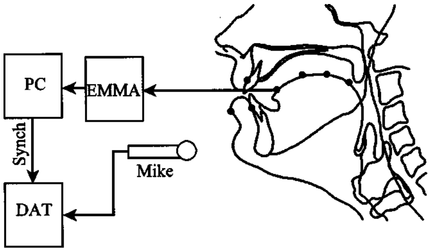

FIG. 1. Experimental setup showing approximate sensor locations (omitting reference sensor on bridge of nose).

The test utterances were formed by inserting the vowels into three different consonant contexts: /p_p/, /t_t/, and /k_k/. These contexts were chosen to give three strong and clearly defined directions of coarticulation with neighboring sounds. Each symmetric CVC sequence, in turn, was embedded in a carrier phrase of the structure *Ich habe geCVCe gesagt* (*I said* _____) with stress on the target vowel. The resulting test words are not lexical items in German but all conform to regular word formation rules. Spellings were devised using regular German spelling rules and the words were presented to the subjects in ordinary German orthography. The speakers read five repetitions of each of the CVC combinations at two different speaking rates, normal and fast.

C. Recordings

Articulatory movements were monitored by means of electromagnetic midsagittal articulography (AG100, Carstens Medizinelektronik). For a general overview of EMMA, see Perkell *et al.* (1992); for an evaluation of the AG100, see Hoole (1996).

In order to register tongue movements, four transducers were mounted on the midline of the tongue at roughly equidistant intervals from about 1 to 6 cm from the tongue tip. The main anatomical reference used was to locate the third coil in line with the rear edge of the lower second molars, with the tongue at rest in the mouth (normally roughly below the junction of the hard and soft palate). Jaw and lower-lip movement were also monitored, but will not be discussed further here. Two coils tracked head movement and were attached to the bridge of the nose and to the border of the upper incisors and gums. Finally, two additional reference coils mounted on a bite-plate were used to define the horizontal axis as the line from the lower edge of the upper central incisors to the lower edge of the upper second molars. Figure 1 shows typical locations of the transducers.

Movements were recorded with a sampling frequency of 250 Hz (low-pass filtered at 35 Hz). The audio signal was recorded on DAT tape, with synchronization pulses generated by the computer on the second channel. For a detailed description of system calibration and data preparation, see Hoole (1996).

D. Experimental procedure

The subjects were tested in two separate recording sessions, usually a few days apart, lasting about 1 h each. In the first recording session the speakers produced the utterances at normal speech rate, in the second recording at a fast speech rate. The consistency of the speech rate across an experimental session was controlled by regular presentations of taped example utterances which were determined for each subject individually in a previous pilot study.

During the recording sessions transducers were monitored with a set of online procedures for evidence of misalignment relative to the transmitter assembly (cf. Perkell *et al.*, 1992; Hoole, 1996).

In a separate session, a reference trace of the midsagittal contour of the subjects' hard palate was made from a dental impression.

For each vowel, one frame of articulatory data was extracted at the acoustically defined midpoint of the vowel. This was generally very close to the point that would be extracted by means of a minimum articulator velocity criterion, but avoided problems with a few systematic cases where minimum velocity was poorly defined, particularly back vowels in /k/-context.

The data were then averaged over the five repetitions of each vowel in each of the three consonantal contexts. This can be expected to improve the fit of the model to the data by removing some random variation and represents a slight departure from the procedure followed in earlier investigations in which individual vowel tokens were analyzed (in radiographic studies multiple repetitions have generally not been available). The use of averaged data appears justified since we are principally interested here in regularities in tongue configuration in the realization of the German vowel system. We have discussed patterns of token-to-token variability in vowel production elsewhere (Hoole and Kühnert, 1995).

After averaging over individual tokens the overall mean of each articulator position was then determined for each subject and subtracted from the data. The data seen by the subsequent algorithm thus consist of displacements from the average articulatory configuration of each subject.

III. ANALYSIS

This main section will trace out the steps required to arrive at a phonetically satisfying model of our dataset. The ride toward a PARAFAC model of vowel articulation turned out to be a bumpy one, and, as already mentioned above, a departure from the basic PARAFAC framework was ultimately required. Identifying in phonetic terms the sources of these difficulties effectively constitutes one of the results of this study.

The procedures followed and the results obtained will be given in the following four subsections: A. Development of the PARAFAC model; B. Discussion of the model; C. Extension of the model; D. Discussion of the extensions.

A. Development of the PARAFAC model

1. A false start

As already mentioned, for PARAFAC analysis it is necessary to choose the number of factors on which to base the model. A preliminary stage therefore involves assessing the number of factors likely to be appropriate to the data. One way of doing this is to apply a (two-mode) principal-component analysis to each speaker individually. If the speaker-specific differences are consistent with the

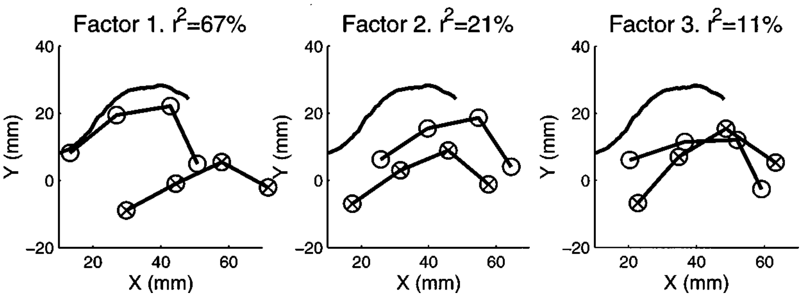

FIG. 2. Example for one speaker of tongue shapes related to the first three components of a principal-component analysis of vowel data. Each panel shows displacement from mean tongue position caused by setting each component in turn to ±2 standard deviations (positive deviation: unfilled circles; negative deviation: circles with crosses). More anterior locations are to the left. Percent variance explained is also indicated.

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 principal-component analysis. For our data, the principal-component analyses consistently gave the picture that three factors captured the data well. The first two factors together usually accounted for about 85% of the variance and generally bore some resemblance to the factors referred to by Harshman *et al.* (1977) as ‘‘front raising’’ and ‘‘back raising,’’ respectively. The third factor, typically accounting for about 12% of the variance, captured the alternation between tongue-blade and tongue-dorsum raising for the (in our data) mutually exclusive consonantal contexts /t/ and /k/. An example of this analysis for one speaker is shown in Fig. 2.

It thus appeared warranted by the data to base the PARAFAC model on three factors. This figure also seemed plausible in phonetic terms, based on the expectation of a roughly two-dimensional vowel space, plus an additional dimension to capture nonvocalic behavior of the tongue-tip. The attempt proved unsuccessful, however. 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 again below to more precise consideration of the properties of the data inconsistent with the PARAFAC model.

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

2. Models for individual consonantal contexts

We present first the results for the vowels spoken in the /p_p/ context, as this can be regarded as the most neutral consonantal context with regard to lingual articulation.

Based on the results of the principal-component analysis and results from the literature mentioned above we 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 (whereas, as a cross-check, a three-factor solution again was not).

Here we should state more explicitly how reliability was assessed. For this stage of the analysis the following two criteria were used: first, the alternating least squares algorithm had to succeed in converging and giving the same solution when initiated from at least six different random starting points; second, acceptable values of a diagnostic for degeneracy had to be obtained. Following Nix *et al.* (in turn quoting Harshman and Lundy, 1984), this was based on the triple product (over the three modes) of the correlations between corresponding sets of weights for each pair of factors (in practice we never have more than one pair). Harshman suggested that triple products more negative than -0.3 are indicative of a degenerate model since the factors in the pair are simply tending to cancel each other out.[1]

For the two-factor solution of the /p/-context material the unexplained variance amounted to 7.7% and the rms error to 1.24 mm. This is very much par for the course: for example, Harshman *et al.* obtained 7.4% variance unexplained and an rms error of 1.74 mm.

As explained in the Introduction, the algorithm provides three sets of weights for each of the two factors: for the articulators (tongue *x* and *y* displacements), for the vowels and for the speakers. After deriving the final PARAFAC model we will look in detail below at the patterns to be observed in each of these sets of weights. Suffice it to say here, for this first analysis based on /p/-context only, that the first factor represents a contrast between high front and low back and the second factor mid front to high back. Particularly for the first factor this is not unlike Harshman *et al.*’s original two-factor solution.

For the /k/-context vowels a two-factor model was also successfully extracted. Both the modelling error (9% variance unexplained, 1.1-mm rms error) and the model itself were very similar to the /p/-context analysis. The latter aspect can be assessed by separately calculating for each factor the triple product of the correlation coefficients between corresponding sets of weights in the /p/-context model and the /k/-context model. Highly similar models would have triple products approaching $+1$.[2] For the two models compared here we obtained values of 0.84 for factor 1 and 0.69 for factor 2 (we note in passing that specifically for the correlation between the vowel weights we would expect a high but not perfect correlation since the two sets of vowels, i.e., those spoken in /p/-context and those spoken in /k/-context, are obviously in some sense different).

Surprisingly, the extraction of a two-factor model for the /t/-context vowels 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—the triple product was strongly negative: −0.56 (the amount of unexplained variance was also rather higher at 13%, although the rms error remained about the same: 1.2 mm). Moreover, the solution was substantially different from the /p/-case, especially for factor 2, the triple product of the correlation coefficients being 0.8 for factor 1 and −0.13 for factor 2.

One possible reason for a degenerate solution can be the extraction of too many factors from the data. At first sight it seems phonetically very implausible that this can be the case here, since it is unclear how one could model a vowel system such as German with just one factor. Indeed, checks made by extracting a one-factor solution for each of the three consonant contexts separately provided no evidence at all that the /t/-context data could be better modeled than the other two contexts with only one factor.

However, as we will see below, there remains a grain of truth in this possibility. A further situation that can lead to degeneracy is inconsistency of the data with the PARAFAC model. As we will also see below, it turns out that the way tongue tip/blade raising is captured by the front two EMMA sensors exhibits speaker-specific patterns that run contrary to the 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 modeling 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 toward a model for the complete data set.

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 to the previous independent analyses of the /p/- and /k/-context vowel material, the unexplained variance and the rms error deteriorated somewhat to 12% and 1.5 mm, respectively. The model for combined /p/- and /k/-context vowel material was very similar to the model extracted for /p/-context only, the triple product of the correlations between combined- and single-consonant models amounting to 0.97 and 0.98 for factors 1 and 2, respectively. The combined-consonant model was also similar to the model extracted for /k/-context only—the corresponding triple products being 0.94 and 0.78.[3]

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 amounted to 20% and the rms error to 1.9 mm. In the subsection below on extending the model we will look in detail at the model error, in particular with regard to subject-specific and subject-independent patterns and with regard to the influence of consonantal coarticulatory effects. But first we will concentrate in the next subsection on discussing in detail 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: triple products of 0.96 and 0.55 for factors 1 and 2, respectively (p-only comparison) and 0.93 and 0.88 (k-only comparison).

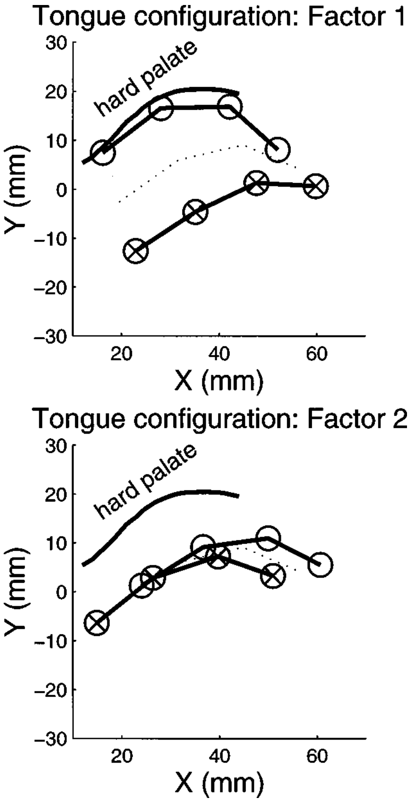

FIG. 3. Tongue shapes related to the factors of the two-factor PARAFAC model of the complete dataset. Each panel shows displacement using mean speaker weights from mean tongue position (shown by dotted line) caused by setting each factor in turn to ±2 standard deviations (positive deviation: unfilled 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.

B. Discussion of the PARAFAC model

Detailed presentation of the two-factor model can proceed most conveniently by taking each of the three sets of weights in turn.

1. Articulator weights

The weights with respect to each factor for the eight articulator coordinates can be shown most vividly by plotting each factor as a pattern of tongue displacement around average tongue position using averaged speaker weights. The result is shown in the two panels of Fig. 3.

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. 3 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. Our second factor is less similar to their second one, however (referred to by them as ‘‘back raising’’). It would share with Harshman *et al.*’s factor 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.

Based on the rationale of Nix *et al.* that there may be advantages in interpretability in analyzing true x/y components of fleshpoint movement rather than displacements along a fixed set of gridlines, one might have expected that our result would be more similar to the Nix *et al.* reanalysis of the Harshman data. But this does not really seem to be the case. Our factor 1 is fairly similar to what (confusingly) emerges as factor 2 in their reanalysis (Nix *et al.*, 1996, Fig. 7b, p. 3715) but their factor does not involve much change in oral opening at the frontmost tongue location. Their factor 1 is similar to our factor 2 in mainly involving retraction versus advancement, but whereas our factor 2 couples slightly higher tongue position with retraction, with them the opposite is the case.

We will return in the concluding discussion to the differences between our solution and other solutions from the literature.

2. Vowel weights

We now turn to consideration of how the German vowel system is represented in the space of the first two factors. The three panels of Fig. 4 show the distribution of the vowels in this space separately for each of the three consonantal contexts.

Factor 1 has been allotted to the ordinate since it has the strongest tongue-raising component; however, since neither factor exclusively involves raising versus lowering, or advancement versus 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 /iː/ and /oː/ for factor 1 and /ɛː/ and /uː/ for factor 2.

Let us first discuss some further 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.[4] 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 (/yː/ closer to /eː/ than to /iː/, etc.). The comparison between unrounded and rounded thus has similarities to that between tense and lax (see also Hoole and Kühnert, 1996). 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 the /t/-context compared to the other two contexts. In /t/-context essentially all vowels except the tense back vowels /uː/ and /oː/ 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 two-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. This is illustrated in terms of the complete fleshpoint data for one vowel of one speaker in the top panel of Fig. 5. The direction of this trend was absolutely consistent over all front vowels and all speakers. In terms of the raw data, the second tongue sensor from the front was located on average about 4 mm more posteriorily in /t/-context than in /k/-context, with generally larger effects for lax vowels than tense vowels and for the normal compared to fast-rate sessions.

A corollary of this finding is that the nominally front vowel /œ/ is located very close to the back vowels /ʊ/ and /ɔ/ in the /t/-context but is widely separated from them in the other two contexts. This is illustrated in terms of the raw fleshpoint data of one speaker in the bottom panel of Fig. 5.

It should be remarked that these strong coarticulatory

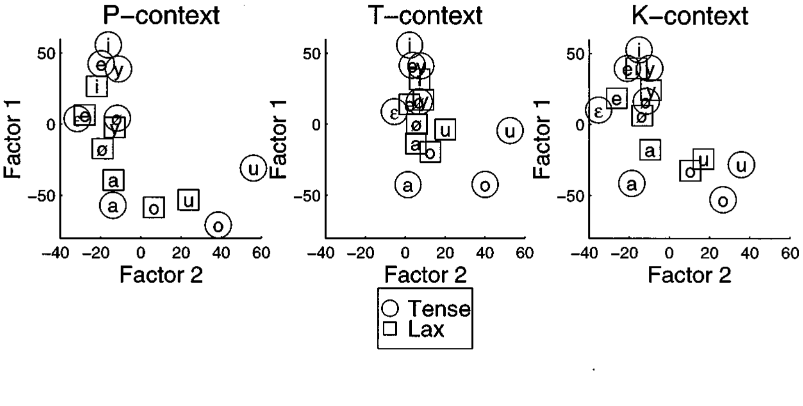

FIG. 4. Distribution of vowels in the factor 1/factor 2 space, shown separately for each of the three consonantal contexts. Lower-case letters i, y, e, ø, a, o, and u are used as generic symbols for the long/short (tense/lax) pairs /iː, ɪ/, /yː, ʏ/, /eː, ɛ/, /øː, œ/, /aː, a/, /oː, ɔ/, and /uː, ʊ/, respectively. The long member of each pair is enclosed in a circle, the short member in a square. ‘‘ɛ’’ with circular enclosure in the figure indicates the long vowel /ɛː/ (no short counterpart).

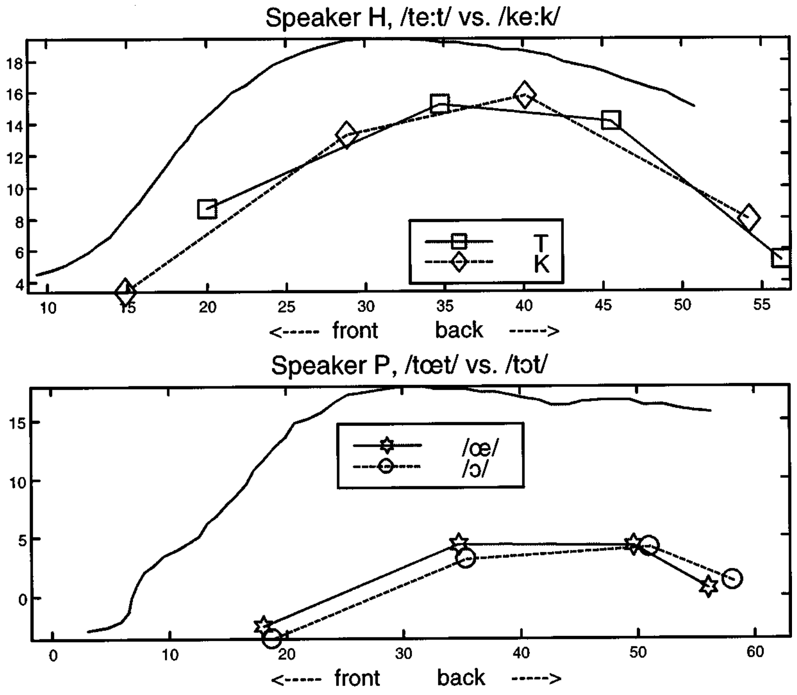

FIG. 5. Two examples of coarticulatory effects on tongue configuration of individual speakers. Top panel: Retraction of front vowel /e:/ in /t/ vs /k/ context. Bottom panel: Approximation of front vowel /œ/ and back vowel /ɔ/ in /t/-context.

effects captured by factor 2 involve advancement/retraction of the complete tongue; it is not until we extend the model below that we will be able to observe more localized coarticulatory contrasts in the region of the tongue-tip—which is where coarticulatory effects of /t/ might, *a priori*, have been expected to be most salient. 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 in the literature. 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 been known for almost 100 years, and has provoked much debate over the course of the century (Meyer, 1910; Wood, 1975; 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.

3. Subject weights

The subject weights are displayed in Fig. 6 with the same assignment of the factors to the *x* and *y* axes as used for the vowel space. From several points of view the pattern of the weights confirms that the extracted model is a satisfactory one. First of all, the sign of the weights is the same for all subjects. If, for a given factor, there were differences in the sign of the subject weights this would indicate that the factor itself is being used to capture subject-specific features. Such a situation would constitute a violation of the modeling assumption of capturing subject-specific effects in a simple scaling of subject-independent factors. Clearly this is not the case in our data.[5]

The relationship of the weights for session 1 (normal rate) versus session 2 (fast rate) is also intuitively satisfying. Essentially one of two patterns occurs: Either the weights for session 2 are located closer to the origin for both factors, indicating a rather straightforward scaling down of articulation (subjects B, C, M, S, and T), or the weights remain close together in the factor 1/factor 2 space, indicating that the subjects made only little change in movement amplitude, but

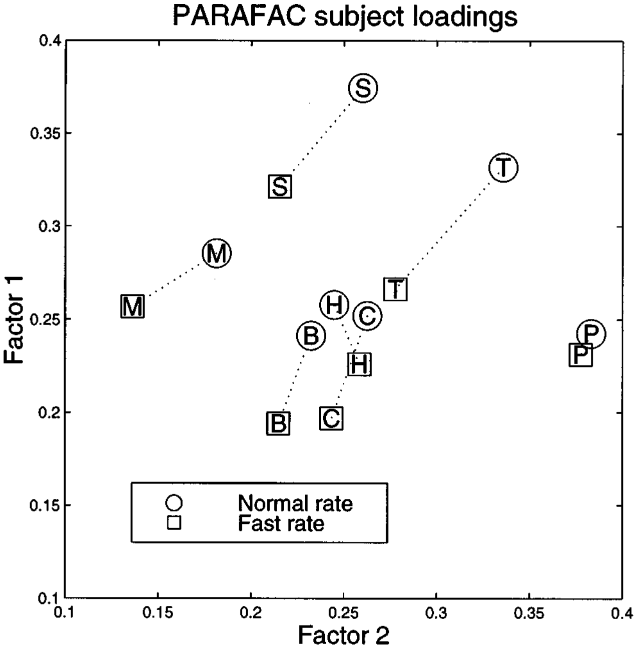

FIG. 6. Distribution of subject weights in the factor 1/factor 2 space. The subject initial is enclosed in a circle for the normal-rate session and in a square for the fast-rate session. The two sessions of each subject are joined by a dashed line.

importantly, remain consistent in their use of these two dimensions of tongue control: this applies to subject P and in slightly less obvious fashion to subject H. Subject P was in fact the subject who made least change in vowel duration over the two sessions. Subject H did have a substantial change in vowel duration; this reflects the simple fact (mentioned in Sec. I) that changes in tempo do not have to be accompanied by a reduction of movement amplitude.

As also discussed in Sec. I, it would have been disturbing if the subjects had shown unsystematic positioning of session 2 weights relative to session 1 weights in the factor space; this would probably have necessitated the conclusion that differences over sessions over which we do not have complete control, such as inevitable slight discrepancies (between subjects, and between the same subject in different sessions) in attachment of the sensors to the tongue, could have seriously deleterious effect on the interpretability of the results. Encouragingly, this does not occur, supporting the validity of this particular modeling approach on the basis of this particular kind of articulatory data.

The patterns in the speaker weights also make clear that a further prerequisite for successful application of the PARAFAC algorithm is met; in order to solve the problem of rotational indeterminacy of the factors, the PARAFAC algorithm requires the presence of differences in the *relative* importance of the factors over subjects (Harshman *et al.*, 1977, p. 699, draw an analogy to the solution of a system of simultaneous equations). Our subject population appears to fulfill this requirement. Although there is a subgroup of subjects quite close to a line with a gradient of 1, indicating fairly restricted differences in the relative importance of the two factors (subjects B, C, H, and T), taking the group as a whole (refer in particular to P, S, and M) a wide range in the relative contribution is covered.

Having made this point, it must be admitted that we now reach the limit of the interpretability of the weights. Thus while we can observe that, for example, speakers S and P represent the two extreme cases, the former making particularly heavy use of factor 1, the latter of factor 2, we can do no more than speculate as to what these differences reflect. Harshman *et al.* consider, for example, whether speaker weights can be related to anatomical features, such as oral cavity length, or relative length of pharynx and oral cavity, but results were rather inconclusive. The main anatomical information we currently have available for our subjects consists of tracings of the hard-palate contour, but these did not provide any obvious clue as to what might lie behind the different relative weight for S and P. Other systematic sources of influence on the speaker weights are certainly conceivable, such as slight differences in accent or in overall articulatory setting (Laver, 1980), but many more speakers would be required to achieve a balanced assessment of these issues.

C. Extending the model

Although the two-factor PARAFAC model discussed in the previous section appeared to give a consistent and revealing picture of vowel articulation per se, the process of extraction pointed to the presence of subject-specific effects, probably related to consonantal coarticulation, not able to be captured in the model. In the present section we outline the approach followed to try to come to terms with these problems and develop a model for the complete data set.

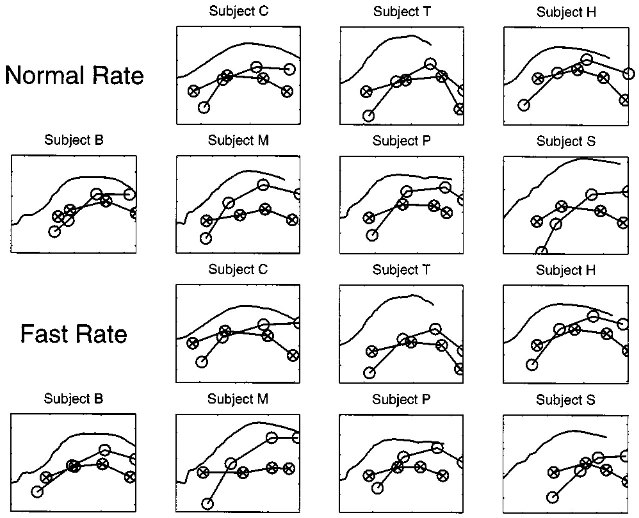

FIG. 7. Patterns of tongue displacement (around mean tongue position) associated with the first principal component of the PARAFAC model error (as in Fig. 3 configuration shown for ±2 standard deviations). Patterns shown separately for each subject and session.

The approach essentially consists of examining the error of the two-factor PARAFAC model for systematic effects. Using the three sets of PARAFAC weights and Eq. (1) we can generate, for each subject separately, the articulatory data predicted by the model. Subtracting this from the original data (the input to the PARAFAC algorithm) gives the model error—for every vowel and for every articulatory coordinate. Note that these subject-specific error matrices still constitute a kind of data that is very similar to the original datasets; whereas these original datasets measure displacement of the tongue from the mean articulator position, the new error datasets measure the displacement of the tongue required to move it from its position predicted by the PARAFAC model to its actual position. Thus following the rationale underlying the whole of this paper, we can now employ procedures such as principal-component analysis to uncover typical pattern(s) of tongue displacement allowing us to succinctly capture a substantial proportion of the variance in the error data.

Carrying out separate principal-component analyses of the error data for each subject showed that the first principal component explained at least 37%, and on average 49%, of the variance in the data. It is revealing to plot the pattern of tongue displacement associated with the first principal component of the analyses. This can be done in a manner entirely analogous to the patterns of tongue displacement shown in Fig. 3 for the factors of the PARAFAC model. However, since there is a separate eigenvector for each subject, this must now be done in a subject-specific fashion. Figure 7 shows the results in this way.

Inspection of Fig. 7 suggests that conceptually this first principal component of the error captures rather similar articulatory behavior in each subject, namely the alternation between tongue-tip/blade raising (presumably for vowels in /t/-context) and tongue-back raising (presumably for vowels in /k/-context). This is fortunate in terms of the aim, to which we still cling, of developing a reasonably parsimonious model of the complete dataset. If these articulatory patterns had shown little in common over subjects, it would have meant that the residuals after extraction of the PARAFAC model consisted of little more than idiosyncratic behavior. Nonetheless, it is important here to point out some clear differences in the patterns over subjects. Because this principal component represents an alternation between tip and back raising, there is an intersection of the tongue contour associated with strongly positive principal-component scores and the tongue contour associated with strongly negative ones. However, the point of intersection differs over subjects. This means that, particularly, the second tongue sensor from the front shows variability in whether it tends to raise or to lower for raised tongue-tip configuration (compare, e.g., subjects P and S, normal rate). The subjects also differ as to whether the raised tongue-tip configuration is associated with fronting (e.g., subject C, normal rate) or retraction (e.g., subject H, normal rate) of the tongue as a whole. These are precisely the kind of subject-specific differences that would be difficult to capture in the PARAFAC framework: It is not obvious how these individual tongue patterns could be generated by means of a simple subject-specific scaling of a subject-independent vector of articulator weights. Presumably we find here the explanation for the failed attempt to derive a three-factor PARAFAC model.

However, we have been assuming that the principal component extracted from the error data is conceptually very similar over subjects—and is readily interpretable phonetically. If this is accepted (and we will demonstrate more formally below that this does indeed seem to be justified) then it suggests that the analysis of consonantal aspects of articulation reacts more sensitively to the precise location of sensors on the tongue (a problem of experimental technique) and/or that the analysis must allow for subjects genuinely differing more as to the precise regions of the tongue involved in the realization of a phonetically defined task (cf. examples in Johnson *et al.*, 1993, p. 701). In other words, consonants are intrinsically less tractable objects for analysis of this kind than vowels (Jackson, 1988, p. 140, discusses why vowels should be particularly *well*-suited to this kind of analysis).

The least parsimonious approach to modeling the complete dataset would now be to retain separately for each subject both the eigenvector (length 8) defining the first principal component of the PARAFAC error, as well as each vowel's score on that component [i.e., 14 subjects × (15 vowels+8 articulator weights)]. Table I shows for each subject the rms error after application of the PARAFAC model (first column) together with the remaining rms error after incorporating the tongue displacements captured by the first principal component of the PARAFAC model error in the completely subject-specific manner just explained (second column). We will refer to the latter as the ideal error.

In the course of deriving the PARAFAC model in Sec. III A above, we suggested that a desirable goal would be to have a model of the complete dataset that has an error magnitude comparable to that found when models are set up for each consonant individually. In those terms our goal would be an rms error of about 1.2 mm. The second column of Table I shows that this aim could be comfortably achieved with a completely subject-specific modeling of the PARAFAC error. But can we still achieve this goal with a more parsimonious approach? Specifically, do we need a separate set of vowel scores on the first principal component of the PARAFAC error for each subject? If we assume—in the spirit of the PARAFAC approach—that this first principal component represents a similar underlying articulatory entity in each subject, and that the subjects employ this entity consistently over speech items, then this should not be necessary. To test this idea we computed a single set of vowel scores by simply averaging over subjects each vowel's score on the first principal component of the PARAFAC error (after first normalizing the scores of each subject to a standard deviation of 1). The resulting rms error is given in the first column of Table II for a model based on the two-factor PARAFAC solution plus averaged vowel scores from principal-component analysis, but with subject-specific eigenvectors from this analysis. The second column of this table gives the amount by which this falls short of the ''ideal'' result given in the second column of Table I.

TABLE I. Column 1: rms error (in mm) for PARAFAC two-factor solution; Column 2: Error after subject-specific principal-component analysis of model error (''ideal'').

Subject	PARAFAC	Ideal
	Normal rate	
C	1.6	1.1
T	1.7	1.0
H	2.0	0.9
B	1.4	0.9
M	2.3	1.5
P	1.9	1.2
S	1.9	1.1
	Fast rate	
C	1.9	1.1
T	1.5	0.9
H	1.7	0.9
B	1.4	1.0
M	2.3	1.2
P	2.2	1.3
S	1.8	1.1
Mean	1.8	1.1

As can be seen, the deterioration in accuracy is very small, amounting to just over 0.1 mm on average. This suggests that there is indeed justification for the PARAFAC-like assumptions just made with respect to this first principal component of the error. The main departure from the PARAFAC conception is the more complex mapping from the underlying articulatory entity to the actual fleshpoint displacements in individual subjects, captured in the subject-specific eigenvectors (articulator weights).

D. Discussion of the extended model

It will be recalled that our initial estimate was that a model consisting of three factors was likely to be appropriate

TABLE II. Column 1: Residual rms error (in mm); Column 2: Shortfall *re*: ideal error given in Table I.

Subject	Residual	Shortfall
	Normal rate	
C	1.2	0.09
T	1.2	0.16
H	1.2	0.28
B	1.1	0.16
M	1.7	0.16
P	1.2	0.04
S	1.3	0.19
	Fast rate	
C	1.1	0.07
T	1.0	0.12
H	1.0	0.07
B	1.0	0.07
M	1.3	0.13
P	1.5	0.14
S	1.2	0.04
Mean	1.2	0.12

to the data. The result of the previous section was in effect to provide the third factor for our model (and we will now refer to it as ‘‘factor 3’’). Articulator and subject aspects of that factor have already been presented in some detail in that section. In the present section we look briefly at the remaining aspect, namely the vowel weights (i.e., the averaged principal-component scores). Since there appeared to be some justification for regarding these weights as an acceptably subject-independent representation, they can be plotted against the vowel weights from the PARAFAC analysis. The most interesting combination seems to be to plot our ‘‘consonantal’’ factor 3 against the second PARAFAC factor, which, as Fig. 4 has shown, incorporated clear contextual effects. This combination is shown in Fig. 8.

Factor 3 on its own separates the speech material with respect to consonantal context quite clearly, with /k/-context material at the positive end of this axis, /t/-context at the negative end, and /p/-context clustering around zero. /t/- and /k/-contexts show almost complete separation (the only exception is that tense /ka:k/ and /tu:t/ items have very similar values, near zero, for factor 3). /k/ and /p/, on the one hand, and /t/ and /p/, on the other hand, show slightly more overlap in terms of factor 3 alone, but taking the factor 2/factor 3 space overall, it is possible to delimit nonoverlapping regions for each consonantal context (/ka:k/ remains the only exception). Factor 3 clearly makes no contribution at all to characterizing individual vowel categories independently of consonantal context. The main regularity with regard to subcategories of vowels is that for /t/- and /k/-context the short lax cognates occupy the more extreme positions on the factor axis. This is very clear for /t/-context, slightly less so for /k/-context where the back vowels are an exception. The more extreme location of the lax vowels is readily understandable in terms of the lax vowels showing stronger consonantal coarticulatory effects at tongue-tip and tongue-back for /t/ and /k/, respectively.

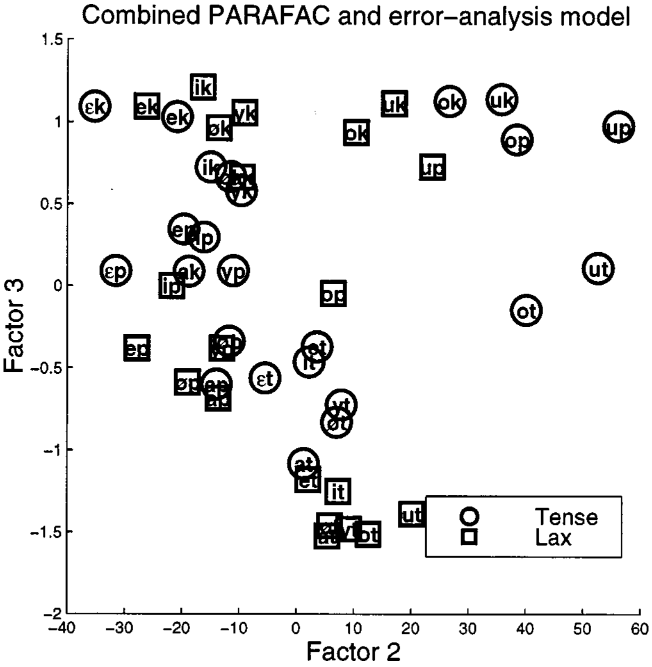

FIG. 8. Distribution of all vowel-consonant combinations in the factor 2/factor 3 space. Factor 2 is the PARAFAC factor also shown in Fig. 4 Factor 3 is derived from the first principal component of the analysis of the PARAFAC error. Same vowel-labeling conventions as in Fig. 4.

IV. GENERAL DISCUSSION

This study confirms on the basis of a particularly large dataset that the control of the tongue for speech is organized around a small number of underlying components. Evidence for this has now accumulated from a number of studies. Investigations restricted to vocalic articulations (e.g., the PARAFAC investigations cited here) have typically extracted two factors;[6] this meshes in well with our study since the two factors from the PARAFAC-based part of our model gave a compact picture of the articulatory structure of the German vowel system. Investigations of both vocalic and consonantal articulation (connected speech) have typically required at least three factors for lingual articulation (e.g., Maeda, 1990; Sanguineti *et al.*, 1998; Badin *et al.*, 1997), one of these obviously being closely related to tongue-tip activity; again this is in close agreement with our result. Thus three components appear to capture much of the shaping of the tongue for speech. Nonetheless, there is a strong possibility that this may underestimate the number slightly. Our present corpus was not intended to capture all aspects of consonantal articulation, but rather to simply maximize the possibility for coarticulatory effects of consonants on vowels for a few important cases (major places of articulation). We will shortly be attempting to apply the techniques developed in this paper to vowel articulations in a richer set of consonantal contexts (and spoken in more natural sentences). The possibility that the number of components should be slightly higher appears all the stronger when it is recalled that our study, as well as many previous ones, has been restricted to mid-sagittal data. Stone and Lundberg (1996) investigated static vowels and consonants using a three-dimensional ultrasound technique and determined by inspection four characteristic ways of shaping the tongue. Yehia and Tiede

(1997) applied a principal-component analysis to vocal tract shapes of static vowels acquired by means of NMRI. Although analysis of the complete vocal tract introduces articulatory features that go considerably beyond the tongue by itself, they found that four basis functions would account for about 90% of the variance in the 3-D shape of the vocal tract, and in fact just taking the first two would already account for about 80% of the variance.

The above discussion relates to the number of components. Regarding the *nature* of the components, in Sec. III B 1 we went briefly into some similarities and differences between our results and earlier investigations. It is clear that in the PARAFAC procedure the nature of the factors into which observable tongue shapes can be decomposed is sensitive to the kind of data used. Nevertheless, the factor referred to by Harshman *et al.* as ‘‘front raising,’’ and emerging as our factor 1, appears to represent a family of tongue shapes that emerges with considerable consistency from very varied investigations.

Regarding our second factor, which mainly involves advancement versus retraction, the fact that such a factor emerged from our analysis (as it does in similar though by no means identical fashion in the ‘‘pseudo-pellet’’ analysis of Nix *et al.*) supports the latter authors’ contention that use of a pre-defined grid system in radiographic analyses may artificially constrain the solutions that can be obtained. Nix *et al.* go further in contending that such a less constrained approach results in a more interpretable solution, specifically in the sense that the orientation of the vowels in the factor space more closely resembles a traditional vowel chart representation. It is indeed noteworthy that the distribution of the American English vowels in the factor space of Nix *et al.* shows very similar values of their factor 2 for front/back pairs such as i/u and e/o that have the same vowel height in traditional descriptions. However, this neat correspondence may be an artifact of the American vowels, which after all are not very cardinal in quality. Their ‘‘pseudo-pellet’’ analysis of the American English vowels involves, by and large, a rotation of the vowel space with respect to the result of the original gridline analysis. However, if a similar rotation is applied to the gridline style analysis they carried out on Jackson’s Icelandic data, then a vowel system with a rather ‘‘nontraditional’’ orientation would result.[7] The Icelandic vowels would actually then be oriented in a manner not dissimilar to the German vowels in our factor space.

In fact, we see no particular problem in accepting that an empirically derived model may appear rotated with respect to traditional representations. Although empirically derived, we believe that our model is eminently interpretable. Indeed, of all the PARAFAC analyses presented to date, we feel that ours fits in most naturally with a plausible pattern of underlying muscle synergies. Specifically, Honda *et al.* (1993) have proposed a two-dimensional physiological space for vowel articulation in which one axis is formed by the agonist–antagonist pairing of Genioglossus Posterior with Hyoglossus, while the second axis is formed by the pairing of Genioglossus Anterior and Styloglossus (see also Maeda and Honda, 1994, especially Fig. 1). These two physiological axes match up very well with our factors 1 and 2, respectively. It is instructive to refer to Fig. 9 of Honda *et al.* (1993) where they display some point vowels in the two-dimensional physiological space defined by the aforementioned axes GGP-HG (ordinate) and GGA-SG (abscissa). In their figure, this coordinate system has been rotated by about 30 degrees, which would align the axes closer to the presumed main line of action of these muscles and in turn brings the physiological vowel space into close coincidence with a traditional vowel space, as they demonstrate by juxtaposing an $F1$ vs $F2$ plot of the vowels. Strikingly, a rotation of our factor space by the same amount (i.e., about 30 degrees) would be quite consistent with the main direction of action of the factors as shown in Fig. 3, and simultaneously would also orient the vowels along more traditional lines in the two-dimensional space (refer back to Fig. 4).

Here we ought to return to the question of the tongue shapes captured by our factors since Maeda and Honda relate the two physiological dimensions GGP-HG and GGA-SG to two of the parameters in the articulatory model that Maeda (1990) had previously derived by factor analysis, namely ‘‘tongue-dorsal-position’’ and ‘‘tongue-dorsal-shape,’’ respectively.

Our factor 1 appears to correspond quite well to the former of these two parameters[8] (involving advancement of the tongue root and raising of the front part of the tongue, cf. Maeda, 1990, Fig. 3b) but our factor 2 seems rather less similar to the Maeda ‘‘tongue-dorsal-shape’’ parameter, which involves a contrast between flat and arched shape. Currently, we are unclear what might explain this difference. Possibly, the different nature of the corpora analyzed may be an influence: Unlike the running-speech corpus of Maeda, our target sounds were restricted to vowels, and, in particular, did not include velar consonants. Functionally, however, Maeda’s parameter and our factor 2 do appear reasonably similar, since they both clearly distinguish high back vowels from front vowels. The amount of variance explained by our factor 2, namely, about 20%, is also strikingly similar to that found by Maeda for his two subjects (23%).

Last of all, we return briefly to the question of vocalic versus consonantal aspects of articulation in this study. We believe that our first two factors are of some general validity, i.e., they capture a complete vowel system in a physiologically plausible manner. With respect to consonant articulation, the study was not aiming for the same level of generalizability. In fact, the important result was simply that these aspects of articulation were less amenable to the highly constrained PARAFAC model of speaker-specific effects. Interestingly, the factor we extracted as factor 3 (essentially by means of a major relaxation of the constraints) did seem compatible with the *spirit* of the PARAFAC approach of a basic articulatory component common to all speakers. Possibly some kind of preliminary transformation of the raw data could have made it feasible to work successfully within the basic PARAFAC framework throughout, even for this consonantal activity. However, this was beyond the scope of the present work, and, by itself, would still not resolve the general question of whether consonantal articulation requires a departure from the PARAFAC approach—simply because our corpus was not designed to give balanced coverage of consonantal articulatory possibilities. This would be an interesting topic for future investigation.

ACKNOWLEDGMENTS

The experiments were performed in collaboration with Barbara Kühnert, Tine Mooshammer, and Christian Kroos, to all of whom thanks for numerous discussions are also due. Thanks also to one anonymous reviewer for incisive comments. For this work we used a MATLAB package developed by R. Bro (available online from http://newton.foodsci.kvl.dk/; see also Bro, 1997). We are grateful to M. Tiede for drawing our attention to this package. Work supported by DFG Ti 69/29-3.

[1] Here we should note in passing that PARAFAC factors are not constrained to be orthogonal, so triple products of exactly zero will not generally occur. Bro (1997) uses a diagnostic referred to as the triple cosine; this is not numerically identical, but is closely related, to the triple product measure used here.

[2] Numerically, the procedure is essentially the same as that outlined above to test for degeneracy, but whereas the degeneracy test considers different factors in the same model, the present test considers the same factors in different models.

[3] In order to be able to correlate the vowel weights, only those vowel weights from the combined model pertaining to the relevant single-consonant model were used, e.g., only p-context vowels from the /pk/-model when comparing to the /p/-only model.

[4] This regularity is least obvious for the pair /øː/ vs /œ/ (both are located very close to 0 on factor 1); however, this difficulty would be resolved if we assume that neutral tongue position actually occurs at slightly negative values of factor 1. This is probably justified since the preponderance of mid to high front vowels in the German system may well displace average tongue position (i.e., zero on the factor axis) away from a neutral position.

[5] Nix *et al.* extracted a three-factor model for English, precisely in order to reinforce the potential for a subject-independent representation in terms of the first two factors; their third factor captured subject-specific behavior of a kind that presumably could not be captured directly in the subject weight matrix.

[6] Jackson extracted three factors for the Icelandic vowel system, but the reanalysis of Nix *et al.* suggested that two factors may actually be more appropriate. Nevertheless, we would not like to exclude the possibility that vowel systems may exist requiring more than two factors.

[7] This seems justified since Nix *et al.* found strikingly similar factors for the English and Icelandic data.

[8] In fact, it might be more correct to say that our factor 1 corresponds to a combination of the ''tongue-dorsum-position'' and ''jaw-position'' parameters in the Maeda model (and which we do not explicitly try to separate in this analysis). In his model these two parameters show considerable similarities. Interestingly, the amount of variance explained by our factor 1—about 57%—is practically identical to the cumulative variance explained by Maeda's jaw- and tongue-position parameters, namely 58% and 56% for his two speakers, respectively.

Badin, P., Baricchi, E., and Vilain, A. (**1997**). ''Determining tongue articulation: From discrete fleshpoints to continuous shadow,'' Proceedings of the European Conference on Speech, Communication, and Technology (Eurospeech 97, Rhodes, Greece), Vol. 1, pp. 47–50.

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

Fischer-Jørgensen, E. (**1985**). ''Some basic vowel features, their articulatory correlates, and their explanatory power in phonology,'' in *Phonetic Linguistics, Essays in Honour of Peter Ladefoged*, edited by V. Fromkin (Academic, New York), pp. 79–99.

Harshman, R., and Lundy, M. (**1984**). ''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. MacDonald (Praeger, New York), pp. 122–215.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). ''Factor analysis of tongue shapes,'' J. Acoust. Soc. Am. **62**, 693–707.

Hashi, M., Westbury, J., and Honda, K. (**1998**). ''Vowel posture normalization,'' J. Acoust. Soc. Am. **104**, 2426–2437.

Honda, K., Hirai, H., and Kusakawa, N. (**1993**). ''Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function,'' Annu. Bull. Res. Inst. Logopedics Phoniatrics, Tokyo **27**, 37–49.

Hoole, P. (**1996**). ''Issues in the acquisition, processing, reduction and parameterization of articulographic data,'' Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation München (FIPKM) **34**, 158–173.

Hoole, P. (**1999**). ''Articulatory-acoustic relations in German vowels,'' Proc. 14th Int. Cong. Phon. Sci., San Francisco (in press).

Hoole, P., and Kroos, C. (**1998**). ''Control of larynx height in vowel production,'' Proc. 5th Int. Conf. Spoken Lang. Processing (Sydney, Australia) **2**, 531–534.

Hoole, P., and Kühnert, B. (**1995**). ''Patterns of lingual variability in German vowel production,'' Proceedings XIIIth Int. Conf. Phon. Sci., Stockholm **2**, 442–445.

Hoole, P., and Kühnert, B. (**1996**). ''Tongue-jaw coordination in German vowel production,'' Proceedings of the 1st ESCA tutorial and research workshop on Speech Production Modelling/4th Speech Production Seminar, Autrans, 1996, pp. 97–100.

Jackson, M. T. T. (**1988**). ''Analysis of tongue positions: Language-specific and cross-linguistic models,'' J. Acoust. Soc. Am. **84**, 124–143.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). ''Individual differences in vowel production,'' J. Acoust. Soc. Am. **94**, 701–715.

Kaburagi, T., and Honda, M. (**1994**). ''Determination of sagittal tongue shape from the positions of points on the tongue surface,'' J. Acoust. Soc. Am. **96**, 1356–1366.

Kroos, C., Hoole, P., Kühnert, B., and Tillmann, H.-G. (**1997**). ''Phonetic evidence for the phonological status of the tense-lax distinction in German,'' Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation München (FIPKM) **35**, 17–26.

Kuehn, D. P., and Moll, K. L. (**1976**). ''A cineradiographic study of VC and CV articulatory velocities,'' J. Phonetics **4**, 303–320.

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge).

Linker, W. (**1982**). ''Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study,'' UCLA Working Papers in Phonetics **56**, 1–134.

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 Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 131–150.

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

Meyer, E. A. (**1910**). ''Untersuchungen über Lautbildung,'' Die neueren Sprachen **18** (Ergänzungsband Festschrift Vietor), 166–248.

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

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). ''Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92**, 3078–3096.

Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.

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

Wood, S. (**1975**). ''The weaknesses of the tongue-arching model of vowel production,'' Phonetics Laboratory, Department of General Linguistics, Lund University, Working Papers **11**, 55–107.

Wood, S. (**1986**). ''The acoustical significance of tongue, lip and larynx maneuvers in rounded palatal vowels,'' J. Acoust. Soc. Am. **80**, 391–401.

Yehia, H., and Tiede, M. (**1997**). ''A parametric three-dimensional model of the vocal-tract based on MRI data,'' Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP97, Munich, Germany) **3**, 1619–1622.