

Arbeitsberichte

36

Institut für

Phonetik und digitale Sprachverarbeitung

Universität

Kiel





**INSTITUTS FÜR PHONETIK UND DIGITALE  
SPRACHVERARBEITUNG**

**UNIVERSITÄT KIEL**

**Arbeitsberichte (AIPUK)**

**Volume 36  
Oktober 2004**



**EDITORS**

**JONATHAN HARRINGTON**

**CHRISTINE MOOSHAMMER**

© 2004 by IPDS Kiel  
All rights reserved.

This report or any part thereof may not be  
reproduced in any form without the written  
permission of

Institut of phonetics and digital speech processing  
University of Kiel

D-24098  
Germany

ISSN 0172 - 8156

# Contents / Inhalt

|  |    |
|--|----|
| <b>On the allophonic behaviour of German /x/ vs /k/ – an EPG investigation</b>           |    |
| Gilbert Ambrazaitis  |    |
| Tina John.....   | 1  |
| <b>Acoustic analysis of phonetic parameters of less masculine sounding German speech</b> |    |
| Karita M. Guzik.....   | 15 |
| <b>Accent in L2 — An acoustic analysis of the vowels of German speakers of English</b>   |    |
| Katrin Liebthal .....  | 30 |
| <b>An EMMA and EPG study on token-to-token variability</b>                               |    |
| Christine Mooshammer   |    |
| Pascal Perrier   |    |
| Susanne Fuchs  |    |
| Christian Geng   |    |
| Daniel Pape .....  | 46 |



# On the allophonic behaviour of German /x/ vs /k/ – an EPG investigation

**Gilbert Ambrazaitis and Tina John<sup>1</sup>**

*Institut für Phonetik und digitale Sprachverarbeitung, Christian-Albrechts-Universität Kiel, Germany*

---

Facts about the phonotactics and the allophonic behaviour of German /x/ vs /k/ are summarized and discussed concerning the question of *categorical vs gradual assimilation*. It is concluded that the different allophonic variation of /x/ vs /k/ can more appropriately be described in terms of a difference in *articulatory control* rather than in *assimilation*. This difference is associated with a stronger allophonic separation for /x/ than for /k/: /x/ (in contrast to /k/) has *allophonic categories* which can be described as different speech sounds as perceived by native speakers. It is hypothesized that these *allophonic categories* are manifested in production through a larger articulatory distance between the most anterior and the most posterior allophone for /x/ than for /k/. The hypothesis is tested successfully by an analysis of /x/ vs /k/ in four subjects using EPG.

---

## 1 Introduction

The allophones of German /x/ and their proposed complementary distribution have been described by Kohler<sup>2</sup> (1995):

“The place of articulation of /x/ varies between palatal and uvular. The palatal fricative occurs after preceding front vowels – including diphthongs with a front target – as well as after the consonants /l, r, n/, and word or morpheme initial; the uvular fricative follows the open back vowels /a(:)/ and /ɔ/; the velar fricative occurs in the remaining cases, i.e. after the back close, close-mid or open-mid vowels including the diphthong with such a target.” (Kohler 1995, our translation)

Thus, following Kohler there are three allophones of /x/: palatal [ç], velar [x] and uvular [χ]. Kohler also describes the allophonic realizations of German /k/:

---

<sup>1</sup> This is a joint paper, and the authors names are listed alphabetically. We blame each other for any errors or other shortcomings in the paper.

<sup>2</sup> Kohler (1995) describes a variety from northern Germany (very close to the standard variety).

“The place of articulation of dorsal stops depends on the following and the preceding vowel and varies between postpalatal and postvelar, depending on the articulatory distance between the two vowels. The more front especially the following vowel is produced, the more palatal is the stop.” (Kohler 1995, our translation)

It has been proposed that the German dorsal fricative /x/ on the one hand and the dorsal stops /k/ and /g/ on the other hand have a different behaviour regarding their allophonic variation. Wiese (1996) remarks:

“It may be the case that the assimilation of the dorsal fricative is categorical, while the dorsal stops (i.e. /g/ and /k/) assimilate in a more gradual and variable manner. If this difference could be verified phonetically, it would provide good evidence for treating DFA [= dorsal fricative assimilation] as a phonological, and the assimilation of /k/ as a phonetic process.” (Wiese 1996)

In this study we understand the term *assimilation* as the process where a segment takes over at least one phonetic feature of an adjacent segment and consequently becomes more similar to it (e.g., Kohler 1995). This definition (which is only one part of Kohler’s definition) does not state whether assimilation is categorical or gradual or whether there are different modes of assimilation, a categorical and a gradual one. Furthermore, it does not state whether it is optional or obligatory. Generally, assimilation has been defined as an optional phenomenon that is dependent on factors like speech rate and style (e.g., Ellis and Hardcastle 2002, Kohler 1995).

Concerning the question of whether place assimilation is categorical or gradual, there are two different accounts. The first one takes place assimilation as a categorical phenomenon which means that the assimilating phonetic parameter is taken over completely (e.g., McCarthy 1988, Spencer 1996). The second account takes place assimilation as a gradual phenomenon and is popular in Laboratory Phonology that uses experimental techniques such as electropalatography (EPG) or electromagnetic articulography (EMA) (see Ellis and Hardcastle 2002 for an overview). The two approaches locate assimilation at different stages of the speech production process: the first one at the abstract planning stage and the second one at the concrete physical execution stage. Hence, in (autosegmental) phonological theory assimilation is a phonological phenomenon and in instrumental research it is a phonetic one. Both accounts generally work with a definition of assimilation similar to the one presented above. In the light of this definition, the segment that assimilates to another segment always belongs to a constituent (e.g., word or morpheme) that has a canonical form, e.g. the German verb /h'a:bən/ where the canonical, unassimilated form is [h'a:bən] and the assimilated form [h'a:bm] (which in spontaneous speech occurs much more frequently than the canonical form). Such a canonical form can be used as a reference point for the decision whether the observed case of assimilation is *categorical* (where the segment changes from one phonological *category* to another) or *gradual* (where the segment does not change the phonological category).

Wiese (1996), however, uses the term *assimilation* in a different way. In the case of the allophones of /x/ or /k/ there are no canonical forms: no words exist that can contain either, e.g., a palatal, or a velar realization of /x/, depending on assimilation factors such as speech

rate or style. Thus, no reference point for classifying an assimilation as *categorical vs gradual* exists in this case. Consequently, we suggest that Wiese's (1996) hypothesis should not be formulated by using the terms *categorical vs gradual assimilation*. However, there still might be a different allophonic behaviour between the fricative and the plosives in that the allophonic differences within fricatives might be categorical but gradient and continuous within plosives. Such a difference, we suggest, is not a difference of assimilation but of *articulatory control*. We have adopted this term from Menzerath and de Lacerda (1933) who for the first described the overlapping articulatory movements by analysing kinematic data and introduced the terms *coarticulation* and *articulatory control* (originally *Koartikulation* and *Steuerung*).

We begin continue by considering some of the salient allophonic properties of /k/ and /x/ in German. It is well known that the left vowel context is relevant for the articulatory control of /x/ (Kohler 1995:81-83). Obviously, the reason for this are phonotactic constraints. /x/ usually occurs after preceding vowels or /l, n, r/ (e.g., /'ɪx/-"I", /'ɛxt/-"real", /d'ax/-"roof", /d'ɔxt/-"wick", /b'u:x/-"book", /m'ɪlx/-"milk", /m'œnx/-"friar", /d'ʊrx/-"through"), but less often before a following vowel. Examples are the diminutive suffix *-xən/* and the plural forms of some of the former examples (/d'ɛxə/-"roofs", /b'yxə/-"books"), as well as words such as /r'axə/-"revenge", /z'axə/-"thing", where the following vowel always (or in most cases) is schwa (ə or ɐ). Exceptions are loan words such as /mazɔx'ɪsmʊs/-"masochism", /xem'i:/- "chemistry", /x'i:na/-"China". In contrast, /k/ can occur before and after vowels (as well as in several consonant clusters that are not relevant here), where in both cases all vowels – and not only schwa – form possible contexts. Examples for /kV/ are /k'i:l/-"Kiel", /k'y:l/-"cool", /k'ɛlə/-"dipper", /k'ø:lə/-"charcoal burner", /k'a:l/-"bald", /k'o:l/-"cabbage", /k'u:lə/-"hollow". Examples for /Vk/ are /n'ɪkən/-"nod", /b'ykən/-"to bend down", /'e:kəl/-"aversion", /ʃt'œkə/-"sticks", /h'a:kən/-"hook", /h'ɔkən/-"to cower", /l'u:kə/-"port hole". So /k/ occurs frequently in both CV and VC contexts, whereas /x/ occurs predominantly in the latter. This in itself might explain why the greatest influence on /x/ is from the left context. Thus, there is a difference in the articulatory control of German /x/ vs /k/ concerning the direction of the control. For /x/ a specific direction of articulatory control is imposed by the phonotactics, while for /k/ the phonotactics do not predict such an orientation towards either the left or the right context. A question here is whether a further difference can be found as well, namely a difference concerning the spacial extension of the articulatory control: does it lead to categorically different allophones in case of /x/, in contrast to /k/?

But what does *categoriality* actually mean regarding allophones of a phoneme? On the one hand, phones can be grouped together into phonological categories, the phonemes – so why should then *allophonic categories* of a phoneme be plausible at all? On the other hand, why should a subcategorization of phonemes not be possible? With such subcategories we mean speech sounds that can be distinguished by native speakers and accepted as *different sounds*, even if – due to their distribution – they are not two different phonemes. We propose that two allophones of /x/, namely the palatal [ç] and a non-palatal, more posterior sound, presumably velar [x] or uvular [χ], are such allophonic categories. In contrast, for /k/ we hypothesize that no such categories can be distinguished, although, of course, contextual variants of /k/ occur, like [k̟] or even [c] in the context of front vowels, and [k̠] in the context of back vowels.

Our aim in this paper is to test whether there are greater articulatory differences in the allophones of /x/ compared with those of /k/ using EPG. Our hypotheses are based largely on Wiese's (1996) auditory impressions and Kohler's (1995) observation that the articulatory distance between the most anterior and the most posterior allophone is larger for /x/ than for

/k/ (for /x/: palatal – uvular; for /k/: postpalatal – postvelar; see above). Following these descriptions and the preceding discussions, we would like to reformulate Wiese’s (1996) hypothesis (see above) as follows:

**Hyp:**

German /x/ and /k/ differ with respect to their allophonic variation, where /x/ has a categorical allophonic variation, while /k/ has not.

- (a) This difference is manifested in perception: /x/ has allophones which are categorically different in the perception of native German speakers, while /k/ has not.
- (b) This difference might have a manifestation in articulation, where the articulatory distance between the most anterior and the most posterior allophone is larger for /x/ than for /k/.

The results of this study will, of course, not provide a satisfactory argument for the entire hypothesis Hyp, since hypothesis (a) still has to be tested by a perceptual investigation. Thus, here we can only test whether Hyp is plausible from an articulatory point of view.

## 2 Method

A remark on the method has to be made here, because two aspects of the method might be problematic for the interpretation of the data. Firstly, Kohler (1990) stresses that there are three allophones of /x/, where the most posterior one is uvular. By using EPG we are, of course, not able to recognize any uvular contacts, and even the velar region might be problematic. So, the question here is whether velar/uvular allophones show EPG contact at all, or – in case they show EPG contact – how these data have to be interpreted (see 4.1).

Secondly, the speech material used in this investigation consists of dorsal fricatives and plosives in a V1-C-V2 structure, where both V1 and V2 are no schwa vowels (2.1). This context is phonotactically unusual for the fricative /x/ in German (see 1). However, in this study we want to test the articulatory control of /x/ vs /k/ due to the vowel context. And since the left vowel context is relevant for /x/, while the right vowel context might be relevant for /k/ (see 1), a VCV context was chosen. Thus, with our data we can also test the observation of the differently directed articulatory control of /x/ vs /k/ and raise the question: is the orientation of /x/ towards the left vowel context strong enough to even work in an unusual context (see 4.2)?

### 2.1 Stimuli

The stimulus set consisted of 120 pairs of invented words. The first word of a pair had the structure /r V C ə n/, where V was one of the following vowels /i ɪ ε a ə u/ and C was either /x/ or /k/ (e.g. /r'ikən/, /r'əxən/). Thus, there were 6 (V) × 2 (/k/,/x/) = 12 words as first word of a pair. The second word of a pair was a hypothetical street name: either /prof'undəve:k/ or /haf'nskive:k/. Each of the 12 possible first words of a pair were combined with the 2 second words of a pair. Thus, there were 24 word pairs. The stimulus set included 2 repetitions of every word pair (= 48 word pairs). This set was replenished with 72 dummy word pairs, that had a similar structure like the word pairs of interest. The 120 word pairs (48 + 72) were randomized and used for the stimulus set of the experiment.

## 2.2 *Speakers*

Four adult male speakers of German aged between 27 and 41 were recorded. All speakers had had previous experience of using EPG palates in experiments. All were native speakers of German but from different areas in Germany. Speaker CGE was born and raised in Baden-Württemberg (south of Germany), speaker DPA in Berlin, speaker JDR and RWI in Mecklenburg-Vorpommern (north of Germany). At the time of the recordings all speakers had been living in Berlin for several years already.

## 2.3 *Data collection*

EPG data were recorded at 100 Hz with the EPG 3 System, which uses an artificial palate with 62 electrodes. The acoustic data were recorded on DAT, digitised at 16 kHz and synchronized with the EPG-signal.

The speakers were seated in a soundproof room in front of a microphone. The randomized 120 stimulus pairs were presented by a computer monitor outside the room through a window. Two separate words, W1 W2, appeared on the screen and the speaker was asked to produce a single word blend out of these words. The blend consisted of the first part of W1 without /ən/ and W2 without the first part of the word (/haf/- or /prof/-). For example: W1 = /r'ɔkən/ W2 = /haf'ɪnskive:k/ resulted in the blend /rɔk'ɪnskive:k/. The final phoneme of W1's first part was always either /k/ or /x/, while the preceding vowel V1 was always one of /i ɪ ε a ɔ u/ and the initial phoneme of W2's remaining part was always one of the vowels V2 /ɪ u/. As a result, the blends included all combinations of {V1 = /i ɪ ε a ɔ u/} × {/x k/} × {V2 = /ɪ u/} (e.g., /ɔkɪ/ in the case of /rɔk'ɪnskive:k/). Speakers were instructed to produce the blend with primary stress on the second syllable (on /kɪn/ in this example), which is also the usual place in which primary stress would fall in German. A test item was presented with instructions prior to running the experiment itself. After the two words appeared on the screen, a tone was sounded with a variable delay between 1-3 seconds which was a cue for the speaker to produce the blend as soon as possible after the tone was heard. Before the actual experiment started, each speaker produced ten practice blends. Trials with errors like hesitating, false starts or technical problems, which were noted during the recording, were immediately presented a second time.

The speech samples were labelled acoustically by means of waveform, spectrogram and f0 contours, which were calculated from the audio signals by the snack-track tool of the EMU software (Cassidy and Harrington 2001). Labelling included marking the acoustic boundaries of the consonant /k/ or /x/ as well as those of the preceding and following vowels. Deviations from the intended realisations of the blend in V1 or V2 were commented. The acoustic boundaries of /k x/ were marked at the onset and offset of the acoustically voiceless interval between the vowels. In the case of /k/ the aspiration was marked separately from the preceding closure.

Analyses were made using R (Cribari-Neto and Zarkos 1999), a software for statistical analysis of data corpora. Articulatory data were extracted over the duration of the consonants. The moments of onset and offset were given by the labelled data. For three speakers, EPG contacts were available even for the context where a uvular realization of /x/ was expected (see 2). Only for one speaker (CGE), realizations of the fricative were found where the EPG signals did not show any contact. These cases probably are uvular realizations – although they were preceded by a close back vowel (and not an open back vowel, see 1). A centre-of-gravity

(COG) parameter was derived from the electropalatographic data over the acoustic interval of /k/ and /x/. COG is defined as in Gibbon & Nicolaidis (1999) as follows:

$$\text{COG} = (\sum R_s F) / \sum R_s$$

where  $R$  is the row number,  $R_s$  the sum of the contacts in a row,  $F$  the weighting factor (with  $F = 7,5$  for row 1 (most anterior) through to  $F = 0,5$  for row 8).

COG measure indicates where tongue-palatal contact occurred. Higher COG values represent more anterior contact whereas lower values represent more posterior contact. The following analyses were made for each consonant /k/ and /x/ separately.

According to the three different vowel contexts that were reported to be relevant for the distribution of the allophones of /x/ (see 1, Kohler 1995) and as a starting point for our analyses, we grouped /x/ according to the preceding vowel contexts “open” ( $V1_o$ ), “back” ( $V1_b$ ) and “front” ( $V1_f$ ), were “front” includes /i ɪ ε/, “open” /a ɔ/ and “back” /ʊ/. For methodological reasons, the same hypothetical categories were chosen for /k/.

### 3 Results

#### 3.1 Influence of V1

We analysed the extent of influence of V1 on /x/ and /k/ close to the acoustic consonant offset, at the 90% time point of the consonant (see Table 1 and Figure 1). For all plosives, the first 90% of the total segment duration was within the closure.

Table 1. Average COG values of /x/ and /k/ per V1-C-V2 context at 90% of segment duration separately for the four speakers

| Speaker<br>C<br>context | CGE   |       | DPA   |       | JDR   |       | RWI   |       |
|-------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
|                         | /x/   | /k/   | /x/   | /k/   | /x/   | /k/   | /x/   | /k/   |
| V1 <sub>f</sub> -C-/ɪ/  | 1.839 | 1.258 | 1.825 | 1.788 | 3.238 | 3.183 | 1.985 | 1.762 |
| V1 <sub>f</sub> -C-/ʊ/  | 1.327 | 0.542 | 0.856 | 0.500 | 2.015 | 0.991 | 1.514 | 0.688 |
| V1 <sub>b</sub> -C-/ɪ/  | 0.667 | 1.218 | 0.700 | 1.471 | 1.300 | 3.218 | 1.364 | 1.498 |
| V1 <sub>b</sub> -C-/ʊ/  | 0.500 | 0.500 | 0.500 | 0.500 | 0.800 | 0.921 | 0.667 | 0.800 |

### *Allophonic behaviour of German /x/ vs /k/*

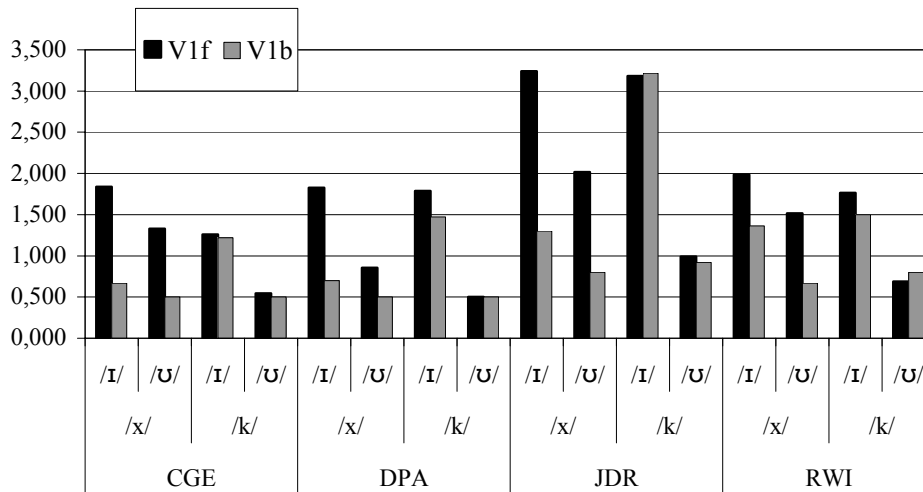


Figure 1. Average COG values of /x/ and /k/ per V1-C-V2 context at 90% of segment duration with V1<sub>f</sub> and V1<sub>b</sub> as left vowel context, /ɪ/ and /ʊ/ as right vowel context shown separately for the four speakers

As Table 1 and Figure 1 show, for all speakers the consonant /x/ shows different COG-values at 90% of segment duration in different contexts, where the highest COG-values appear in V1<sub>f</sub>-/x/-/ɪ/, lower values in V1<sub>f</sub>-/x/-/ʊ/, a further step lower in V1<sub>b</sub>-/x/-/ɪ/ and lowest values in V1<sub>b</sub>-/x/-/ʊ/. Data for /k/ show for all speakers that there are always much lower COG-values when the consonant is followed by /ʊ/ than by /ɪ/. The COG values for /k/ only differ little in respect to the left vowel contexts V1<sub>f</sub> and V1<sub>b</sub> in the same right vowel context. For all speakers except CGE values for /k/ and /x/ in the V1<sub>f</sub>-C-/ɪ/ context are very close, where COG-values in the V1<sub>b</sub>-/k/-/ɪ/ context are close to these, too.

### *3.2 Allophonic variation: /x/ compared with /k/*

COG values were calculated on EPG data extracted at the acoustic midpoint of /x/ and /k/ and averaged according to all 3 (V1 = "front"/"back"/"open") × 2 (V2 = /ɪ ʊ/) = 6 possible left and right context combinations separately per consonant. The results were plotted by fitting normal distribution curves to the data (Figure 2). The figure gives a first impression of how the different vowel contexts influence the place of articulation of the consonants at its temporal midpoint.

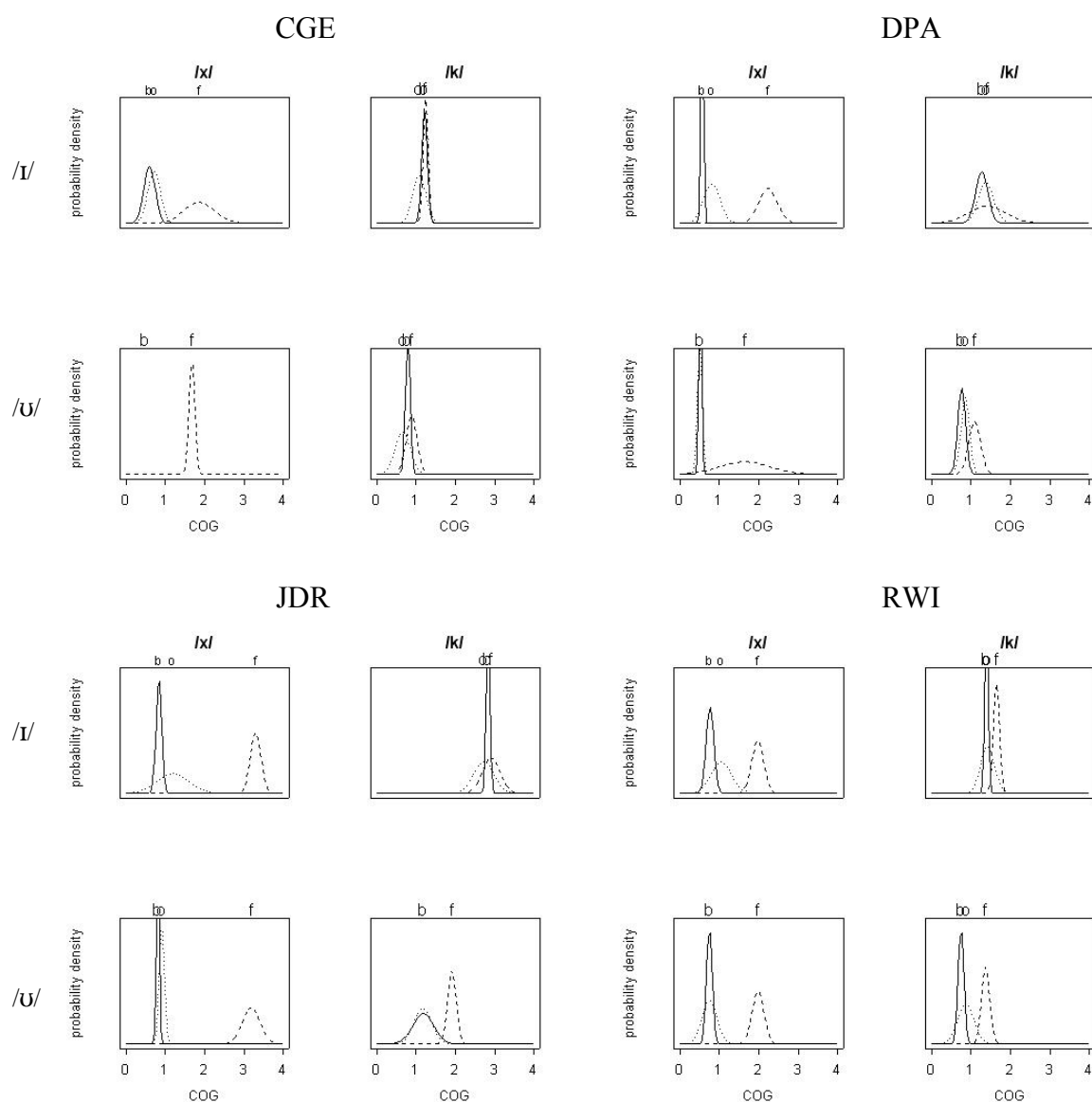


Figure 2. Normal distributions of COG values for /x/ and /k/ at 50% of consonant duration in the contexts “V1-C-V2” per V1 categories f = “front” (dashed curves) vs o = “open” (dotted curves) vs b = “back” (solid curves) with V2 /i/ in the first line and /u/ in the second line for all of the four speakers separately.

Generally it can be observed from Figure 2 that there are differences in the normal distribution curves of the COG-values for /x/ vs /k/, while the differences of the curves comparing V2 /i/ vs /u/ per consonant are smaller than the differences of /x/ vs /k/ in the same left and right vowel contexts. The figures for /x/ show clearer separations between the curves than the figures for /k/. How many divisions can be made depends on the speaker. One division in the V1-/x/-/i/ context can probably be made for all the four speakers. Also in the V1-/x/-/u/ context only one division can be made for the speakers DPA, JDR and RWI and none for CGE (because of the missing data). No clear separations between the curves for different vowel contexts are shown in the figures for /k/ except for speaker JDR in the V1-/k/-/u/ context where one can divide two areas. Though it is possible that the data of different

speakers can be separated into the same number of divisions one has to consider that the points where these separations are made differ from speaker to speaker. The separation point between the categories within V1-/x/-/u/ is close to COG = 2 for JDR but COG = 1.5 for RWI. Also when no separation is possible the range of data depends on the speaker. The COG value range is 0.5 - 1.5 for CGE but 0.5 - 2.5 for DPA in the V1-/k/-/i/ context.

In Figure 2, none of the figures for /x/ show clear separations for the three hypothetical categories (“front”, “back”, “open”). Thus, the place of articulation of the allophones of /x/ – as expressed by COG values – more likely indicates an articulatory manifestation of two than of three allophonic categories. For /k/, the overlap of these categories is much larger than for /x/.

Table 2. Average Bayesian distances (bdist) of COG-values for /x/ at 50% of consonant duration between the contexts V1<sub>o</sub>-/x/-V2 V1<sub>f</sub>-/x/-V2 (V1<sub>o</sub> to V1<sub>f</sub>) and V1<sub>o</sub>-/x/-V2 V1<sub>b</sub>-/x/-V2 (V1<sub>o</sub> to V1<sub>b</sub>) per V2 /u/ and /i/. Significance of differences of the Bayesian distances between V1<sub>o</sub> to V1<sub>f</sub> and V1<sub>o</sub> to V1<sub>b</sub> is given in the last row of the table. The absolute values of the given numbers represent the distances (the minus sign only adds information about the direction).

| Speaker<br>V2                      | CGE    |         | DPA     |         | JDR     |         | RWI    |     |
|------------------------------------|--------|---------|---------|---------|---------|---------|--------|-----|
|                                    | /u/    | /i/     | /u/     | /i/     | /u/     | /i/     | /u/    | /i/ |
| bdist                              |        |         |         |         |         |         |        |     |
| V1 <sub>o</sub> to V1 <sub>f</sub> | -4.595 | -14.787 | -11.352 | -36.858 | -46.073 | -12.578 | -8.636 |     |
| V1 <sub>o</sub> to V1 <sub>b</sub> | 1.556  | 2.943   | -0.493  | 2.440   | -5.285  | 2.362   | -2.239 |     |
| p-value                            | 0.008  | 0.008   | 0.008   | 0.004   | 0.004   | 0.008   | 0.109  |     |
|                                    | *      | *       | *       | **      | **      | *       | n.s.   |     |

In order to quantify the extent of overlap between the normal curves in Figure 2, we calculated the Bayesian distances (bdist) between them. If a pair of normal curves overlaps completely, then bdist is zero and the more the normal curves are separated (= the less likely the two distributions belong to a single category), the greater is bdist. Assuming from Figure 2 that the place of articulation of the consonant following the open vowel would be between the place of articulation after front and back vowels (see 4.1 for a discussion of this aspect), we calculated bdist between V1<sub>o</sub>-/x/-V2 and V1<sub>f</sub>-/x/-V2 as well as between V1<sub>o</sub>-/x/-V2 and V1<sub>b</sub>-/x/-V2 separately in the following V2 = /i/ and V2 = /u/ contexts (so 4 calculations for /x/). The Bayesian distance between V1<sub>o</sub>-/x/-V2 and V1<sub>f</sub>-/x/-V2 was tested whether it differs significantly from the Bayesian distance between V1<sub>o</sub>-/x/-V2 and V1<sub>b</sub>-/x/-V2 using a Wilcoxon test. We then did the same but for the /k/ context (so another 4 bdist calculations, hence 8 in total for /x/ and /k/ together). We calculated these 8 bdist values separately for each of the 4 speakers. Distances and the results of the Wilcoxon tests are given in Table 2 for /x/ and in Table 3 for /k/.

For all speakers except RWI the Bayesian distance of the COG-values between /x/ preceded by open vowels and /x/ preceded by vowels of the front category is significantly larger than the distance between /x/ preceded by open vowels and /x/ preceded by the back vowel, independently of the following vowel /i/ or /u/ (see Table 2). The corresponding distances for /k/ are significantly different for DPA and JDR when followed by /u/ and RWI when the consonant precedes /i/ (see Table 3).

Table 3. Average Bayesian distance (bdist) of COG-values for /k/ at 50% of consonant duration between the contexts open V1<sub>o</sub>-/k/-V2 V1<sub>f</sub>-/k/-V2 (V1<sub>o</sub> to V1<sub>f</sub>) and V1<sub>o</sub>-/k/-V2 V1<sub>b</sub>-/k/-V2 (V1<sub>o</sub> to V1<sub>b</sub>) per V2 /ʊ/ and /ɪ/. Significance of differences of the Bayesian distances between V1<sub>o</sub> to V1<sub>f</sub> and V1<sub>o</sub> to V1<sub>b</sub> is given in the last row of the table. The absolute values of the given numbers represent the distances (the minus sign only adds information about the direction).

| Speaker<br>V2<br>bdist             | CGE   |       | DPA   |        | JDR   |       | RWI    |     |
|------------------------------------|-------|-------|-------|--------|-------|-------|--------|-----|
|                                    | /ʊ/   | /ɪ/   | /ʊ/   | /ɪ/    | /ʊ/   | /ɪ/   | /ʊ/    | /ɪ/ |
| V1 <sub>o</sub> to V1 <sub>f</sub> | 3,467 | 1,378 | 0,651 | -60,18 | 0,819 | 0,762 | 1,604  |     |
| V1 <sub>o</sub> to V1 <sub>b</sub> | 3,466 | 1,856 | 1,631 | -6,39  | 1,036 | 1,271 | -2,483 |     |
| p-value                            | 0,932 | 0,046 | 0,799 | 0,021  | 0,283 | 0,442 | 0,004  |     |
|                                    | n.s   | *     | n.s   | *      | n.s   | n.s   | **     |     |

Further analyses disregard “open” as third category and add it to the ”back“ category except for speaker RWI-data with V2 = /ɪ/.

For the main analysis with respect to the hypothesis (b) the Bayesian distance of the overlaps of the normal distribution curves of the COG-values between the contexts V1<sub>f</sub> and V1<sub>b</sub> (where V1<sub>b</sub> includes V1<sub>o</sub>, see above) were calculated for each consonant separately for V2. The differences between the distance for /x/ and /k/ were judged statistically. Results are shown in Table 4.

Table 4. Average Bayesian distance (bdist) of COG-values at 50% of consonant duration between the contexts V1<sub>f</sub>-C-V2 and V1<sub>b</sub>-C-V2 where V1<sub>b</sub> includes V1<sub>o</sub> (except for RWI with V2 = /ɪ/) for C /x/ and /k/ per V2 /ʊ/ and /ɪ/. Significance of differences of the Bayesian distances between V1<sub>f</sub>-C-V2 V1<sub>b</sub>-C-V2 for C /x/ vs /k/ is given in the last row of the table.

| Speaker<br>V2<br>C | CGE     |         | DPA     |         | JDR     |         | RWI     |     |
|--------------------|---------|---------|---------|---------|---------|---------|---------|-----|
|                    | /ʊ/     | /ɪ/     | /ʊ/     | /ɪ/     | /ʊ/     | /ɪ/     | /ʊ/     | /ɪ/ |
| /x/                | -5.539  | -15.002 | -9.158  | -45.719 | -24.558 | -8.648  | -11.913 |     |
| /k/                | 2.020   | 1.363   | 0.694   | -2.017  | 1.252   | -0.399  | 1.635   |     |
| p-value            | 0.00004 | 0.00006 | 0.00004 | 0.00004 | 0.00004 | 0.00004 | 0.0002  |     |
|                    | ***     | ***     | ***     | ***     | ***     | ***     | ***     |     |

For all speakers the distances of COG-values between the two hypothetical allophonic categories “front” and “back” are significantly higher for /x/ than for /k/.

## 4 Discussion

### 4.1 Uvular allophones of /x/

Regarding the question about uvular allophones of /x/ and the suitability of the EPG System to measure this place of articulation (see 2) our analyses show quite confusing results. First of all, the artificial palate does not cover the entire section of the soft palate. Thus, for both consonants that were expected to be velar or uvular a potential difference between the actual articulation and the one reflected by the EPG data has to be considered. It is possible (and likely) that for these consonants only the anterior part of the constriction area is captured by the EPG recording, rather than the whole constriction area – or even that no contact at all is recorded, in case the constriction is too posterior.

Consequently, it should actually not be surprising that for one speaker (CGE) we did not measure any EPG contact in contexts where a velar allophone of /x/ was expected (see 3.1). But on the other hand, it is quite surprising that for those realizations of /x/ that were expected as uvulars, EPG data *could* be recorded. Furthermore, for all speakers (except CGE), the COG values of those consonants that were preceded by an open vowel tended to be *between* the values of the other two hypothetical categories (Figure 2). In other words: those consonants that were expected to be more posterior (uvulars in the case of /x/) were articulated more anterior than those expected to be velars – at least as far as the articulation is represented by the COG values.

How should such strikingly anterior values of expected uvulars compared to expected velars be interpreted? One possible explanation could simply be the allophonic contrast of /x/ as realized by (at least three of) our speakers: maybe they do not make any distinction between the allophones that were described as velar vs uvular (Kohler 1995, see 1) and they prefer the velar one. In the case of CGE, uvular allophones seem to occur but in a slightly different context than the one described by Kohler (1995). Both observations could possibly be explained by the dialectal influence of the speakers (see 2.2): none of them speaks exactly the same variety which Kohler (1995) describes, and CGE is the only speaker with a southern German accent.

### 4.2 The direction of articulatory control of /x/ vs /k/

The first analysis (3.1) showed that near the segment offset of /x/ (at 90% of consonant duration) the left vowel still has more influence on the consonant's place of articulation than the (temporally much closer) right vowel, while for /k/ there was more coarticulation with the right vowel context. For /k/ all four speakers always produced higher COG-values where the right context was /ɪ/ than where the right context was /ʊ/ – independently of the left vowel context. These results (Table 1, Figure 1) show clearly that the right vowel context has a coarticulatory influence on both /k/ and /x/, while the left vowel context has a huge influence on the place of articulation of /x/ and hardly any influence at all on that of /k/. Thus, although an unusual context for /x/ was chosen for the stimuli (V1-C-V2, with V1, V2: not ə, ɐ), stronger influence of the left context on /x/ was clearly apparent.

Such unusual context could lead to exceptions from our results as a reflection of comparable “real” exceptions that happen with real words in German. One example is the

word /mazɔx'ismus/-“masochism” where at least two different pronunciations exist in German: /x/ can be pronounced as palatal or velar/uvular. If the pronunciation is derived from the usual rule, i.e. “the left vowel determines the place of articulation”, then the pronunciation will more likely be non-palatal. However, if it is derived from the morphology in an “incorrect” way, the pronunciation can be palatal: There are speakers who seem to interpret /x'is(mus)/ as a morpheme and consequently /x/ as morpheme initially which leads to the realization as a palatal fricative (although the first morpheme of the word is the name /mazɔx/). This problem probably occurs more likely on unknown words or words with a very low frequency in every day speech. The problem of unknown or unfrequent words in some cases could even lead to totally irregular realizations of /x/ (where, in contrast, the variation in the “masochism” example still can be treated as regular). In fact, we commented such irregularities for speaker DPA during the labelling process (see 2.3). For some reason, these realisations did not affect our general results.

#### 4.3 Differences in place of articulation: allophones of /x/ vs allophones of /k/

Our results have also shown that there might be quite a lot of variability in the precise places of articulation for the allophones of /k/ and /x/ – at least as far as they are represented by the COG values – due to the individual speaker as well as due to the right vowel context. For none of the speakers there seemed to be three distinct allophonic categories of /x/, thus only two allophones are found to be more likely. By contrast, there was no such distinct categorical allophonic variation for /k/. However, this was the case for only three speakers. Speaker JDR had noticeably fronter realisations of /k/ when the preceding vowel was front than when it was back.

The results illustrated in Table 2 (3.2) confirm that there are more likely to be two allophones of /x/ than three, since the Bayesian distances between the distributions of the  $V1_f$ -/x/-V2 context and that of the  $V1_o$ -/x/-V2 context were significantly larger than the distances between the  $V1_o$ -/x/-V2 and the  $V1_b$ -/x/-V2 contexts, at least for most of the tested cases (speaker RWI with V2 = /ɪ/ was the only exception). The results for /k/ (Table 3) also showed significant differences in the tested distances, although it seems very unlikely from Figure 2 that there are different allophones for /k/ with one exceptions (JDR with V2 = /ʊ/).

In sum, the results for /k/ indicate that there are no clearly distinct allophonic groupings. With regard to the results for /x/ we found that the distinction between a preceding open and a preceding back vowel produced no serious allophonic difference. We have also shown that the articulatory distance between the most anterior allophones and the most posterior allophones is significantly larger for /x/ than for /k/. That is, from an articulatory point of view it is much more likely for /x/ than for /k/ that there are two distinct allophonic categories.

## 5 Conclusion: General results and speaker variability

In summary, the hypothesis (b) (see 1) was confirmed by the results of this study: the articulatory distance between the posterior and the anterior allophones is significantly larger for /x/ than for /k/. Thus, in case our hypothesis of categorical allophonic variation of /x/ (as defined in 1) can be confirmed by further perceptual studies, the actual study can provide additional articulatory evidence.

As far as our general result is concerned, we observed less speaker dependent variability than expected. But even when speaker variability does occur, it mostly supports this conclusion. In the case of CGE, who seemed most likely to produce uvular allophones, this exception would lead to an even larger contrast of posterior vs anterior allophones than the one found in this study. Other examples of speaker variability in this study did not contradict the general results either, but reflect differences in the range of articulations. E.g., the precise place of articulation (as expressed by COG values) varies a lot between speakers (see 4.3).

## 6 Future research

It follows that hypothesis (a) (see 1) should be tested by further perceptual investigations. This indeed might seem superfluous to a German speaking phonetician, since the ability of German listeners to distinguish [ç] from [x / χ] on the one hand, and the inability to (easily) distinguish [k̟] vs [k̠] as well as [x] vs [χ] on the other hand seems obvious from every day observation. Or, how Kohler puts it concerning /x/: “for German speakers and listeners, the opposition between the front and back ends of the front-to-back continuum is a psychological reality: they can produce and perceive the distinction without difficulty in any context” (Kohler 1990:46). Thus – although perceptual studies are, of course, possible – we would like to propose another very simple production experiment, in order to test this “psychological reality”. Speakers of German could be confronted with an orthographic stimulus, where a consonant is underlined. The speakers could be asked to produce the sound of the underlined consonant of the presented word (this should be demonstrated by the researcher once, e.g. for the stimulus “Harke” (/h'arkə/-“rake”) the researcher produces the underlined consonant as [k<sup>h</sup>ə]). The stimuli here could contain the following words:

- (a) “Küche” (/k'yxə/-“kitchen”),
- (b) “Buch” (/b'u:x/-“book”),
- (c) “Dach” (/d'ax/-“roof”),
- (d) “Kino” (/k'i:no/-“cinema”),
- (e) “Kuhle” (/k'u:lə/-“hollow”),
- (f) “Kanu” (/k'a:nu/-“canoe”),

where the fricative is preceded by a front vowel (a), a back closed vowel (b) and an open vowel (c), and the stops in (d-f) are followed by corresponding vowels. We would expect the subjects to recognize the same consonant category for (d-f) and consequently produce a “neutral” [k<sup>h</sup>ə]. We would further expect subjects to recognize the fricative sound in (a) as a different consonant category as in (b-c) and consequently produce a palatal [ç] in (a), while the production in (b) should be [x] or [χ] and not differ significantly from (c).

Our articulatory data showed that it is very likely that a palatal category of /x/ is separated from the non-palatal realizations while it remained unclear whether a contrast between “velar” and “uvular” exists, or at least, whether the one described by Kohler (1990, 1995) is dialect-specific. Thus, further studies should concentrate on these non-palatal variants. Here, it might be useful to apply a different instrumental technique such as electromagnetic articulography.

We have seen that the left context exerts a considerable influence on the articulatory control of /x/ while near the segment offset only the right context had an influence on the place of articulation of /k/. However, we have not checked whether the right context influences /k/ to the same extent as the left one influences /x/.

For future experiments similar to the one reported here speakers will be more carefully selected according to regional accent and we also plan to record a greater number of tokens.

The investigation was part of an undergraduate course supervised by Jonathan Harrington. We planned and carried out the whole experiment in association with the other three participants of the course, Lasse Bombien, Gesche Brauer and Annika Grunwald who also labelled most of the data. The data collection was made possible by the staff at ZAS, Berlin, especially Jörg Dreyer who managed the whole recording session and, of course, by the four subjects. Klaus Kohler read an earlier version of the paper and provided us with very helpful comments concerning some theoretical issues. The actual version benefits a lot from his advice. Christine Mooshammer and Jonathan Harrington read and commented on this reformulated version and had some further helpful suggestions. To all these people we express our sincere gratitude.

## References

- Cassidy, S. & Harrington, J. (2001): Multi-level annotation in the Emu speech database management system. *Speech Communication* 33, 61-77.
- Cribari-Neto, F. & Zarkos, S. G. (1999): R: Yet another econometric programming environment. *Journal of Applied Econometrics* 14, 319-329.
- Ellis, L. & Hardcastle, W.J. (2002): Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics* 30, 373-396.
- Gibbon, F. & Nicolaidis, K. (1999): Palatography. In: Hardcastle, W.J. & Hewlett, N. (eds): *Coarticulation in Speech Production: Theory, Data, and Techniques*. Cambridge: Cambridge University Press; pp. 229-245.
- Kohler, K. J. (1990): Comment on German. *Journal of the International Phonetic Association* 20(2), 44-46.
- Kohler, K. J. (1995)<sup>2</sup>: *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- McCarthy, J. J. (1988): Feature Geometry and Dependency. *Phonetica* 43, 84-108.
- Menzerath, P. & de Lacerda, A. (1933): *Koartikulation, Steuerung und Lautabgrenzung*. Berlin, Bonn: Ferd. Dümmers Verlag.
- Spencer, A. (1996): *Phonology: Theory and Description*. Oxford: Blackwell.
- Wiese, R. (1996): *The Phonology of German*. The Phonology of the worlds languages. Oxford: Clarendon Press. 7-218 and 271.

# Acoustic analysis of phonetic parameters of less masculine sounding German speech

**Karita M. Guzik**

*Institut für Phonetik und digitale Sprachverarbeitung, Christian–Albrechts–Universität  
Kiel, Germany*

---

The acoustic characteristics of two speakers were explored, one less masculine sounding (LMS) and one more masculine sounding (MMS), as judged by a group of listeners. The basic questions are whether the LMS speaker differs significantly from the MMS speaker in the first and second formant values as well as mean fundamental frequency and pitch range, and whether these differences are similar to those found between male and female speech. The investigation shows that significant differences are found within the vowels (the LMS speaker having less centralized vowels) as well as in fundamental frequency (the LMS speaker having a higher median value and higher quartile values). Anatomical, acoustic and sociophonetic reasons for such differences are explored. The results suggest that there is a tendency for the LMS speaker to approximate towards female speech characteristics.

---

## 1. Introduction

Male and female voices contain many acoustically coded differences which make it possible for the listener to identify the speaker's sex. But there are also cases when female speakers are labelled as masculine sounding, or male speakers as feminine sounding. There are only a few studies of the phonetic characteristics of these speaker groups, and by contrast to male-sounding female speakers, the phonetic characteristics of female-sounding male speakers seem to be more phonetically distinct (Moonwomon, 1985). Furthermore, almost all studies of less masculine sounding male speech are restricted to English.

The main aim of this study is to examine some acoustic characteristics of less masculine sounding speech in German, and also to investigate whether these characteristics show similarities to the phonetic attributes of female speech.

Avery and Liss (1996) characterize "less masculine sounding" speech as being located somewhere in the middle of the continuum from female to male speech, where gender-specific attributes may be assigned incorrectly. There might be a misidentification of gender, or alternatively a mismatch of "biological" and "social" gender (Biemans and van Bezooijen, 1999): the listener associates a biologically male speaker with social female qualities (or vice versa).

There is general agreement that male and female speech differs in a number of ways. As has been demonstrated in many studies (e.g., Fant, 1975; Henton, 1995; Simpson, 1998; Diehl *et al.*, 1996), the formant values of male and female vowels differ within the same vowel category such that the female F1 × F2 vowel space has been found to be consistently larger than the male. The relationship between male and female formant values is different for the first and second formant as well as for each vowel category. This non-uniform relationship has been observed for a number of languages (see Henton, 1995; Fant, 1975). Because of these differences it is not possible to derive the formant values of female vowels from male vowels of the same category by using a single scale factor.

As Simpson (2001) states, different reasons have been proposed to account for this non-uniformity: according to Fant, (1966, 1975) the differences found in male and female vowel spaces are related to anatomical differences between male and female vocal tracts. The formant values are reciprocally proportional to vocal tract length (Fant, 1960). As this relationship does not account for the observed non-uniformity, separate scale factors have to be calculated for each formant and vowel category.

Simpson (2001) took a biomechanical approach: due to the anatomical differences in male and female vocal tract dimensions, the articulators move through different distances in male and female speech. Simpson's results show that women cover these distances with greater articulatory speed and so are able to produce more peripheral vowels than men over the same duration. In Simpson's analysis, then, non-uniformity arises from the different distances through which the articulators move in producing the different vowel categories.

During vowel production, there is a complex interaction between fundamental frequency and formants. Due to women's higher fundamental frequency, female harmonics are further apart. Diehl *et al.* (1996) and Goldstein (1980) state that distances between vowel categories have to be increased to maintain the same degree of audible distinctivity at a high fundamental frequency. This could suggest that the greater female vowel space may result from the need to compensate for the higher fundamental frequency and correspondingly poorer differentiation of the vowels.

The factor of voice quality is also of importance. Titze (1989), Henton and Bladon (1985), Holmberg *et al.* (1988) and Klatt and Klatt (1990) find consistent voice quality differences between male and female speech: women tend to use breathy voice more often than men. Titze sees the reason for this voice quality in anatomical and physiological differences of the vocal cords, whereas Henton and Bladon (1985) argue in sociophonetic terms.

Some of the differences between male and female speech arise from gender-typical behaviour (Labov, 1972, 1990). As Henton (1995) argues, women tend to hyperarticulate: they produce vowels with a wider mouth opening. Henton states that such open-mouth articulation can be induced by the need to articulate clearly, in particular in passing on language to their children.

Average female fundamental frequency, due to biological differences such as shorter and thinner vocal cords, is about one octave higher than male (Titze, 1989). Perception

experiments carried out by Coleman (1976) suggest a very high correlation between the position of the fundamental frequency and the degree of masculinity or femininity in the voice.

Contrary to the assumption that women use a wider pitch range Snidecor indicates as early as 1951 that female variability is rather smaller. Henton re-examines pitch ranges from different sources (Peterson and Barney, 1952; Fitch and Holbrook, 1970; Fant, 1973; Hudson and Holbrook, 1981; Graddol and Swann, 1983) and points out that most of them have investigated variability on the linear Hertz scale, where women seem to have a greater pitch range. If a logarithmic scale is used, as Henton does both for the re-examined sources and in a further investigation, no significant difference between male and female pitch ranges is observed. Henton argues that male speakers decrease and female speakers increase their variability intentionally if they want to enhance gender-typical behaviour.

As previously stated, most investigations into less masculine sounding speech are restricted to English. In all studies there is agreement that listeners have a clear idea of the perceptual impression of less masculine sounding speech.

Avery and Liss (1996) investigated fundamental frequency contours, vowel formant midpoint values, spectral moments of fricatives and the speaking rate of four “less masculine sounding” and four “more masculine sounding” speakers, as judged by a group of listeners in a paired-comparison perceptual test. The results revealed that the less masculine sounding speakers used a significantly greater extent of downward shifts with steeper slopes, and they also produced a rising intonation pattern more frequently. Avery and Liss further observed that the F1 and F2 values for the vowels /i/, /æ/ and /A/ were higher for the less masculine sounding than for the more masculine sounding speakers, so that their vowel quadrilateral was displaced upwards and to the right compared to that of the more masculine sounding speakers. The less masculine sounding speakers produced less centralized vowels in connected speech as well as in a /hVd/ context, though a significant effect could only be observed for the second formant of the vowel /i/. Less masculine sounding speakers produced higher /s/ centre frequencies and less diffuse /j/ energy distribution than the more masculine sounding speakers. According to Avery and Liss, this arises from increased or reduced lip rounding, indicating that hyperarticulation took place during the production of these fricatives. Finally, the less masculine sounding speakers used a slightly but not significantly higher speech rate.

Avery and Liss suggest two possible reasons for the observed differences between the two speaker groups. First, the differences could be caused by different stylistic behaviour (as the hyperarticulation within the fricatives suggests) and second, the differences in vowel production could arise from different anatomical conditions.

Terango (1966) reports that the “effeminate-sounding” speakers in his experiment used a higher fundamental frequency, but this value was lower than the mean value reported for American men as a whole. Wolfe *et al.* (1990), (cited by Gaudio, 1994), reported higher fundamental frequency values for their “feminine sounding” speakers, but a greater degree of femininity could not be achieved through an increase in fundamental frequency alone. Gaudio (1994) as well as Smyth *et al.* (2003) observed no influence of

fundamental frequency on the rating as “gay/straight-sounding”. But while repeating the perceptual experiment with the categories “masculine sounding” and “feminine sounding” Gaudio (1994) reports that a higher mean fundamental frequency led to the reassessment as “feminine sounding” of several speakers previously rated as “straight-sounding” and Smyth *et al.* (2003) found that listeners were reluctant to rate a gay-sounding voice as “feminine” if it had a low fundamental frequency.

As far as pitch range is concerned, Crystal (1975) found that “effeminate-sounding” speakers use greater variability and greater dynamic range, whereas according to Terango (1966) neither speaker group showed significant differences. Gaudio (1994) himself observed a relation between greater pitch range and the rating of the speakers as “effeminate-sounding”. Additionally, Wolfe *et al.* (1990) state that “feminine sounding” speakers use a greater number of rising patterns. This is consistent with the findings of Avery and Liss (1996). Gaudio (1994) argues that pitch range can have an influence on the rating as “effeminate-sounding” only in interaction with other factors such as intonation patterns. Smyth *et al.* (2003) examine the effects of different discourse types and different listener groups on the rating of male voices within an continuum from “very gay sounding” to “very straight sounding”. Also, Smyth *et al.* (2003) state that speakers use different numbers of parameters to different degrees in order to be judged as “gay-sounding”.

The reasons for all these reported differences are not yet known. According to Smyth *et al.* (2003), “gay-sounding” speakers practise imitation of female speech behaviour, although some of the parameters used by speakers are not those that differentiate male and female speech. Gaudio (1994) argues that the terms “feminine/effeminate-sounding” are inadequate, as there is no shift in the direction of female speech. But he does not deny that, within the perception, there is an observable correlation between these terms.

The aim in this paper is to present an acoustic analysis of the stressed German vowels /i:/, /e:/, /a:/, /o:/ and /u:/ as well as mean and standard deviation of fundamental frequency. Based on the results of the cited studies for English I hypothesize the following:

1. Listeners show clear consistency in judging LMS and MMS speakers.
2. The vowel formant values of LMS speakers differ from those of MMS speakers.
3. This deviation is not equal for different vowel categories.
4. LMS speakers produce less centralized vowels.
5. The mean fundamental frequency of LMS speakers is higher than for MMS speakers.
6. The variability of fundamental frequency is greater for LMS speakers than for MMS speakers.

| Vowel            | /i:/   | /e:/  | /a:/                                     | /o:/    | /u:/       |
|------------------|--|---|--|---------|------------|
| Word             | Berlin<br>sieben<br>ziemlich<br>ließ<br>viel | Lebensmittel<br>lesen<br>jeder<br>ebenfalls<br>stehen<br>kam<br>bat | Laden<br>Abend<br>Ladentür<br>kam<br>kam | gedroht | Wut<br>gut |
| Number of tokens | 5  | 5   | 7  | 1       | 2          |

Table 1: The vowel types and the words from which they were extracted in *Die Buttergeschichte*.

## 2. Method

### 2.1. Materials and recordings

The speech data used in this experiment included recordings from one LMS and from one MMS speaker as well as speech data from four speakers from the *Kiel Corpus of German Read Speech* (IPDS 1994), for which recordings were already available. All speakers came from the Schleswig-Holstein area and were between 24 and 31 years old. The speech data for all six speakers consisted of the story *Die Buttergeschichte* and the relevant vowels are shown in Table 1.

The speech data for all six speakers were recorded under studio conditions at the laboratory of the Institute for Phonetics and Digital Speech Processing, University of Kiel. The speakers were instructed to use a normal speech rate and a normal level of speaking loudness while reading the story. The recorded samples were digitized at a sampling rate of 16 kHz.

### 2.2. Perceptual task

A preliminary listening experiment was carried out in which 3 male and 23 female subjects were asked to rank the speakers on a scale of perceived masculinity. A five-point scale was used between 1 (most masculine sounding) and 5 (least masculine sounding). For all six speakers, a single sentence was extracted from the *Buttergeschichte*. A warning tone was prepended to each sentence and the sentences were presented to the listeners one at a time in random order over a loudspeaker. After each sentence, the subjects were asked to rate the speaker in terms of his masculinity on the 5 point scale. The subjects were given just over 5 seconds to make each response.

### 2.3. Analysis

The acoustic analysis is based on the entire speech material provided by the two speakers judged as LMS and MMS. The speech samples of both speakers were analysed with the speech signal processing program *xassp* (IPDS, 1997). Manual segmentation and labelling of the speech data of both speakers was carried out on the basis of waveform, spectrogram and auditory impression. For subsequent analysis the *EMU system for speech database analysis* (Cassidy and Harrington, 2001) was used.

All stressed vowels of the categories /i:/, /e:/, /a:/, /o:/ and /u:/ were extracted. The first four formant frequencies derived automatically using an LPC analysis program (Scheffers and Simpson, 1995; Simpson, 1998). The first and second formant frequencies were extracted at the temporal midpoints of the vowels.

The fundamental frequency distribution was extracted over the voiced frames of the whole speech material for each speaker using an algorithm by Schäfer-Vincent, (1982, 1983) included in *xassp*. Median and quartile values were calculated using *R*.

Since Traummüller und Erikson (1995) have argued that gender-specific behaviour can occur in perception the results of male and female listeners were evaluated separately.

## 3. Results

### 3.1. Listening experiment

Table 2 shows that 82.6% of female and 83.3% of male listeners judged the MMS speaker (Speaker 1) as more masculine sounding (1 on the scale); 93.5% of female and 83.3% of male listeners the LMS speaker (Speaker 6) as less masculine sounding (5 on the scale). There is therefore strong support from these results for distinguishing between these two speakers LMS and MMS is the way originally hypothesised.

### 3.2. Vowel quality

Table 3 shows mean values and standard deviation for the F1 and F2 values of the vowels /i:/, /e:/ and /a:/ as measured at the temporal midpoint. As there are too few tokens for the vowels /o:/ and /u:/, mean values and standard deviation could not be calculated. Table 3 shows single values instead. The mean values and distribution for all vowels in the F1×F2 space is shown in Figure 1.

The mean F2 values for /i:/ and /e:/ are 2449 Hz and 2476 Hz respectively for the LMS speaker; and 1987 Hz and 1946 Hz respectively for the MMS speaker. This shows a great difference between the speakers of about 500 Hz, whereas for the F1 values, the difference is only 47 Hz for /i:/ and 92 Hz for /e:/. For the vowel /a:/ the LMS speaker also produces higher F1 values: at 902 Hz the value is 297 Hz higher than the corresponding value for the MMS speaker and in the second formant the LMS speaker is 391 Hz higher.

| Female listeners |       |       |       |       |      |       |
|------------------|-------|-------|-------|-------|------|-------|
|                  | Scale |       |       |       |      |       |
| Speaker          | 1     | 2     | 3     | 4     | 5    | Total |
| 1                | 82.6  | 15.2  | 2.2   | 0     | 0    | 100   |
| 2                | 19.5  | 43.5  | 37    | 0     | 0    | 100   |
| 3                | 13    | 60.9  | 23.9  | 2.2   | 0    | 100   |
| 4                | 19.5  | 34.8  | 39.1  | 4.4   | 2.2  | 100   |
| 5                | 0     | 32.6  | 39.1  | 28.3  | 0    | 100   |
| 6                | 0     | 0     | 0     | 6.5   | 93.5 | 100   |
| Male listeners   |       |       |       |       |      |       |
|                  | Scale |       |       |       |      |       |
| Speaker          | 1     | 2     | 3     | 4     | 5    | Total |
| 1                | 83.3  | 16.7  | 0     | 0     | 0    | 100   |
| 2                | 0     | 33.33 | 33.33 | 33.33 | 0    | 100   |
| 3                | 0     | 33.33 | 66.7  | 0     | 0    | 100   |
| 4                | 16.7  | 50    | 33.33 | 0     | 0    | 100   |
| 5                | 0     | 0     | 66.7  | 33.33 | 0    | 100   |
| 6                | 0     | 0     | 0     | 16.7  | 83.3 | 100   |

Table 2: Percentage of female (above) and male (below) listeners that gave the speakers a category rating on a five point scale from 1 (most) to 5 (least) masculine sounding.

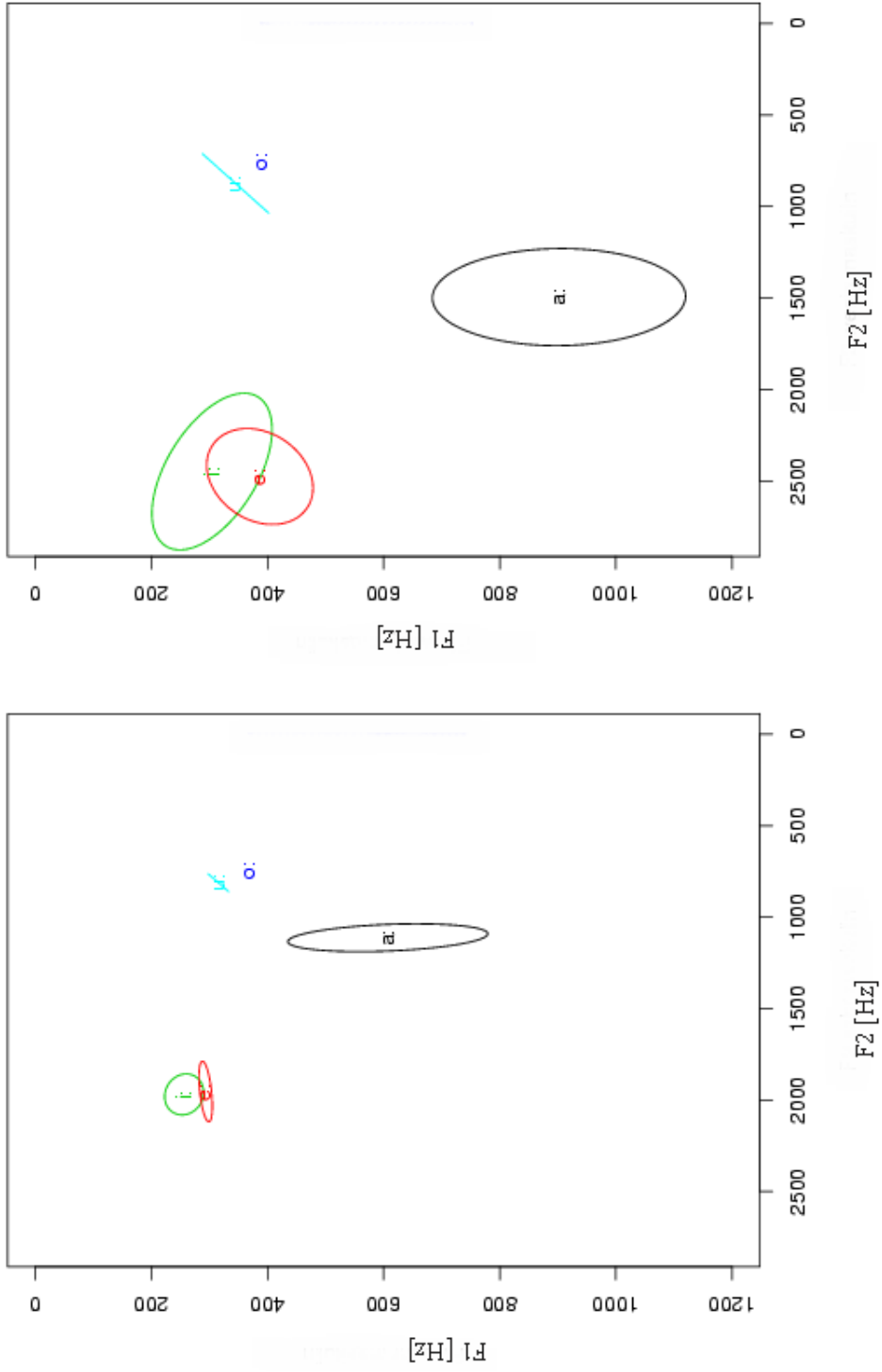


Figure 1: Standard deviation ellipses (which include approximately 95% of the data points) for the vowels /i:/, /e:/, /a:/, /o:/ and /u:/ of the MMS (left) and the LMS (right) speaker in the  $F1 \times F2$  space using the Hertz scale.

Looking at the standard deviation of the LMS speaker for the F1 and F2 values it is clear that it is in all cases greater than that of the MMS speaker.

For the one token of /o:/ the LMS speaker produces an F1 value of 390 Hz and an F2 value of 754 Hz whereas the MMS speaker has his F1 value at 367 Hz and his F2 value at 720 Hz. For the vowel /u:/ the LMS speaker produces one higher and one lower F1 value (402 Hz compared to 289 Hz for the MMS speaker and 288 Hz compared to 333 Hz for the MMS speaker) as well as one higher and one lower F2 value (1036 Hz compared to 784 Hz for the MMS speaker and 712 Hz compared to 740 Hz for the MMS speaker).

In order to verify the hypothesis that the LMS speaker produced more peripheral

| Speaker | Vowel | Formant | Mean [Hz] | Standard dev. [Hz] | n |
|---------|-------|---------|-----------|--------------------|---|
| MMS     | /i:/  | F1      | 257       | 14                 | 5 |
|         |       | F2      | 1987      | 67                 |   |
| LMS     | /i:/  | F1      | 304       | 42                 | 5 |
|         |       | F2      | 2449      | 175                |   |
| MMS     | /e:/  | F1      | 295       | 4                  | 5 |
|         |       | F2      | 1946      | 70                 |   |
| LMS     | /e:/  | F1      | 387       | 37                 | 5 |
|         |       | F2      | 2476      | 107                |   |
| MMS     | /a:/  | F1      | 605       | 70                 | 7 |
|         |       | F2      | 1104      | 33                 |   |
| LMS     | /a:/  | F1      | 902       | 89                 | 7 |
|         |       | F2      | 1495      | 108                |   |
| Speaker | Vowel | Word    | Formant   | Value[Hz]          | n |
| MMS     | /o:/  | gedroht | F1        | 367                | 1 |
|         |       |         | F2        | 720                |   |
| LMS     | /o:/  | gedroht | F1        | 390                | 1 |
|         |       |         | F2        | 754                |   |
| MMS     | /u:/  | Wut     | F1        | 333                | 1 |
|         |       |         | F2        | 740                |   |
| LMS     | /u:/  | Wut     | F1        | 288                | 1 |
|         |       |         | F2        | 712                |   |
| MMS     | /u:/  | gut     | F1        | 298                | 1 |
|         |       |         | F2        | 784                |   |
| LMS     | /u:/  | gut     | F1        | 402                | 1 |
|         |       |         | F2        | 1036               |   |

Table 3: Mean and standard deviation for F1 and F2 values of the vowels /i:/, /e:/, /a:/ as well as F1 and F2 values of the vowels /o:/ and /u:/ for both speakers. All values were extracted from the temporal midpoints of the vowels.

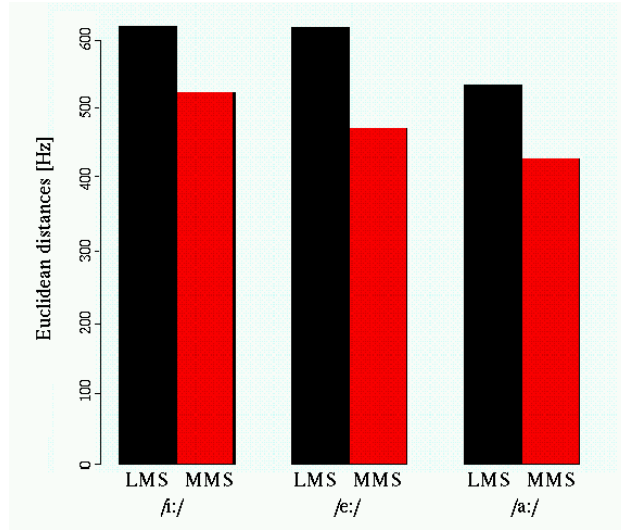


Figure 2: Mean Euclidean distances to the centroid /i:/, /e:/ and /a:/ for the LMS (black) and the MMS (grey) speakers.

vowels, the Euclidean distances were calculated for each vowel token to the average F1 x F2 across all vowels separately for each speaker. (The back vowels /u:/ and /o:/ were not included in this Euclidean distance calculation because of the small number of tokens). As Wright (2003) has recently shown, the Euclidean distance to the centre of the vowel space is a good measure for the extent of peripherality of vowels in the formant space.

$$E_k = \sqrt{(F1_k - \overline{F1})^2 + (F2_k - \overline{F2})^2} \quad (1)$$

The results, which are summarized in the bar plot in Figure 2, indicate that the LMS speaker has greater Euclidean distances than the MMS speaker. The LMS speaker also shows greater Euclidean distances for /i:/ and /e:/ than for /a:/. Overall, LMS had significantly greater Euclidean distances than MMS to the centre of the F1 x F2 vowel space ( $t=2.60$ ,  $df=29.12$ ,  $p < 0.05$ ).

### 3.3. Fundamental frequency

Figure 3 shows the distribution of fundamental frequency for both speakers. As the distribution is non-symmetric, the median and quartile values (rather than mean and standard deviation) had to be calculated.

The median of the LMS speaker (143 Hz) was clearly higher than the corresponding value for the MMS speaker (91 Hz). The quartile distance of the LMS speaker was also higher, at 49 Hz. An extraordinarily high maximum of the LMS speaker at 435 Hz was observed <sup>1</sup>, whereas the MMS speaker had his highest value at 145 Hz.

<sup>1</sup>which was no error in the F0-algorithm.

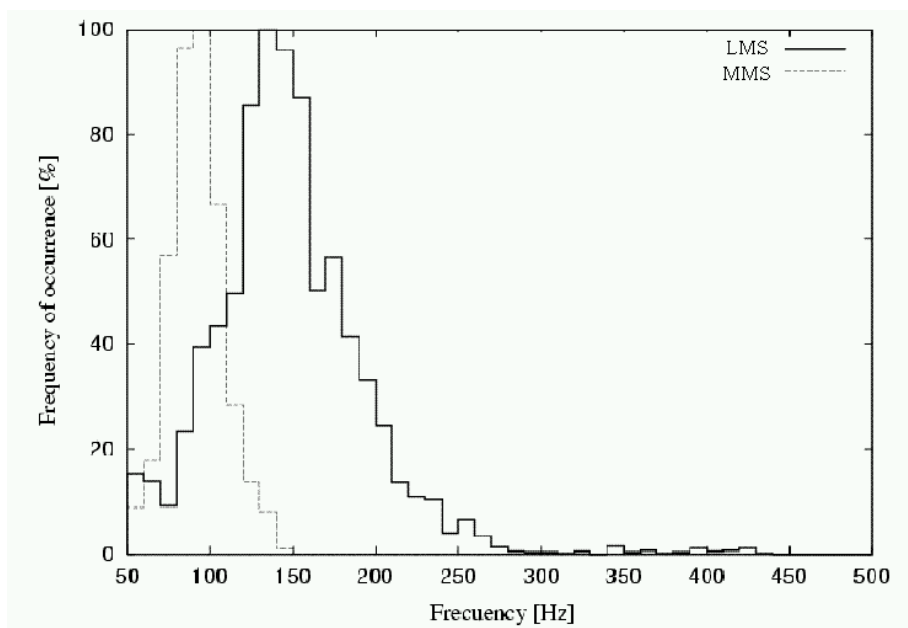


Figure 3: F0 distributions of the LMS and MMS speakers.

A Wilcoxon rank sum test with continuity correction confirmed the observable differences: the distributions of the two speakers were significantly different ( $W = 15754856$ ,  $p < 0.0001$ ).

The results of this study can be summarized as follows:

- The vowels of the LMS speaker in the acoustic  $F1 \times F2$  vowel space are significantly more peripheral than the corresponding vowels of the MMS speaker.
- The variation between the two speakers is different for each vowel category.
- The median value of the LMS speaker is clearly higher than that of the MMS speaker.
- The quartile values of the LMS speaker are higher than those of the MMS speaker on the linear as well as on the logarithmic frequency scale.

#### 4. Discussion

The first experiment showed that German listeners were able to differentiate reliably less, from more masculine sounding speakers. This finding is consistent with results from previously reported studies (Avery and Liss, 1996). But the question arises as to which phonetic parameters play the key role in this perception. Gaudio (1994) has suggested that there might be an interaction between several cues (e.g., pitch range and intonation patterns), but the relations between these cues are not yet known. This could suggest that if one cue is missing, another one is more distinct, so that trading relations take place.

The acoustic analysis showed that the LMS speaker produces less centralized vowels than the MMS speaker, which is compatible with the results of Avery and Liss (1996). It was seen that the production of peripheral vowels is not similarly distinct for all vowel categories. This non-uniformity recalls that found in comparing male and female speech, and could indeed suggest an approximation of the LMS speaker towards female speech. Several reasons could account for the differences between the two speakers.

First, the different formant values of the vowels could be attributable to a difference in vocal tract dimensions. Though there were no obvious differences in the speakers' heights and body size, we can't discount the possibility that the LMS speaker has a smaller vocal tract. To verify this possibility it would be necessary to examine anatomical data.

Second, the higher fundamental frequency of the LMS speaker could also have contributed to his more peripheral vowels. To maintain the same degree of audible distinctivity at a high fundamental frequency, the distances between neighbouring vowel categories must be increased (Diehl *et al.*, 1996).

Third, voice quality, as Simpson (2001) states, could also play an important role. Though there was no explicit analysis of voice quality in this experiment, auditory impressions suggest that the LMS speaker tends to use a breathier voice quality. This tendency has also been demonstrated for female speech (Titze, 1989; Henton and Bladon, 1985).

Fourth, according to Smyth *et al.* (2003), the phonetic behaviour of "gay-sounding" speakers could be caused by the attempt to approximate their speech stylistically towards typical female speech. As Henton (1995) states, vowel hyperarticulation is a typical attribute of female linguistic behaviour. So the production of less centralized vowels could be regarded as an effort on the part of the LMS speaker to imitate female speech behaviour.

As far as fundamental frequency is concerned, the LMS speaker shows a greater median value and greater quartile values than the MMS speaker. Avery and Liss (1996) found no differences between their speaker groups whereas Terango (1966) states that his "effeminate-sounding" speakers had a higher mean fundamental frequency than other speakers.

Anatomical reasons notwithstanding, the statistically significant higher mean fundamental frequency of the LMS speaker in this experiment could be induced by an attempt on the part of the speaker to achieve a higher degree of femininity in his voice. According to Coleman (1976), there is a correlation between perceived femininity and high funda-

mental frequency. But Wolfe *et al.* (1990) on the other hand emphasize that a greater degree of femininity could not be achieved through an increase of the fundamental frequency alone.

The greater F<sub>0</sub>-variability of the LMS speaker on the linear as well as the logarithmic scale could also be sociophonetically motivated, as women who want to sound more feminine use a greater pitch range (Henton, 1989). This sociophonetic approach would give credence to the view that the LMS speaker tries to copy female linguistic behaviour.

Though parallels were observed between phonetic characteristics of female and less masculine sounding speech, it is too soon to say whether a shift towards a more female-sounding voice is the only thing that occurs when men speak in a less masculine sounding way. As Smyth *et al.* (2003) emphasize, “gay-sounding” speakers show variations in areas which do not mark distinctive differences between male and female speech. To obtain more information about these areas more experiments are needed, addressing more phonetic parameters such as consonantal production strategies, durational structures and intonation patterns. Additionally, the investigation of the use of different voice qualities has been neglected so far, but this could also be characteristic of a less masculine sounding speech style in German more than in English. Indeed, the LMS speaker tends to use breathy voice, like typical female speakers, but furthermore there is evidence of an added use of nasalization. So articulatory experiments on vocal cord vibration and velum movement are needed here. Moreover, data from transgender speakers could also shed new light on the issues raised in this paper, particularly where it comes to the adaptation of typical voice attributes of the other gender. And finally, comparisons with other languages are indispensable and clearly needed here, since variants that are expressive in one language could be phonemic in another.

I am very grateful to Jonathan Harrington, Katrin Liebthal, Alexander Radtke, Jonathan Rodgers, Michael Scheffers, Adrian Simpson, Thomas Wesener and Michael L. Willson for their comments and support.

## References

- Avery, J. D. and J. M. Liss (1996). Acoustic characteristics of less-masculine-sounding male speech. *Journal of the Acoustical Society of America* 99, 3738–3748.
- Biemans, M. and R. van Bezooijen (1999). Biological gender and social gender in relation to voice quality. In *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, USA, 1–7 August 1999*, pp. 1249–1252.
- Cassidy, S. and J. Harrington (2001). Multi-level annotation in the EMU speech database management system. *Speech Communication* 33, 61–77.
- Coleman, R. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, 168–180.
- Crystal, D. (1975). *The English Tone of Voice: Essays in Intonation, Prosody and Paralanguage*. London: Arnold.

- Diehl, R. L., B. Lindblom, K. A. Hoemeke, and R. P. Fahey (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24, 187–208.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *STL-QPSR (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm)* 4, 22–30.
- Fant, G. (1973). Stops in CV syllables. In G. Fant (Ed.), *Speech Sounds and Features*, pp. 110–139. Cambridge, MA: MIT Press.
- Fant, G. (1975). Non-uniform vowel normalization. *STL-QPSR (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm)* 2–3, 1–19.
- Fitch, J. L. and A. Holbrook (1970). Modal fundamental frequency of young adults. *Archives of Otolaryngology* 92, 379–382.
- Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech* 69(1), 30–57.
- Goldstein, U. (1980). *An articulatory model for the vocal tracts of growing children*. Ph. D. thesis, M.I.T., Cambridge, MA.
- Graddol, D. and J. Swann (1983). Speaking fundamental frequency: some physical and social correlates. *Language and Speech* 26, 351–366.
- Henton, C. and R. Bladon (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, 221–227.
- Henton, G. (1995). Cross-language variation in the vowels of female and male speakers. In *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, 13–19 August, 1995*, Volume 4, Stockholm, pp. 420–423.
- Holmberg, E., R. Hillmann, and J. Perkell (1988). Glottal airflow and pressure measurements for soft, normal and loud voice by male and female speakers. *Journal of the Acoustical Society of America* 84, 511–529.
- Hudson, A. I. and A. Holbrook (1981). A study of the reading fundamental vocal frequency of young Black adults. *Journal of Speech and Hearing Research* 24, 197–201.
- IPDS (1994). *The Kiel Corpus of Read Speech*, Volume 1, CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1997). *xassp User's Manual (Advanced Speech Signal Processor under the X Window System)*. In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 31–115.
- Klatt, D. H. and L. C. Klatt (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820–857.
- Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2, 205–254.
- Moonwomon, B. (1985). Towards the study of lesbian speech. In *Proceedings of the First Berkely Women and Language Conference*, Berkely, pp. 96–107. Woman and Language Group.

- Peterson, G. E. and H. L. Barney (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175–184.
- Schäfer-Vincent, K. (1982). Significant points: Pitch period detection as a problem of segmentation. *Phonetica* 39, 241–253.
- Schäfer-Vincent, K. (1983). Pitch period detection and chaining: Method and evaluation. *Phonetica* 40, 177–202.
- Scheffers, M. T. M. and A. P. Simpson (1995). LACS: Label assisted copy synthesis. In *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, 13–19 August, 1995*, Volume 2, Stockholm, pp. 346–349.
- Simpson, A. P. (1998). *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*. AIPUK 33.
- Simpson, Adrian, P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *Journal of the Acoustical Society of America*, 2153–2164.
- Smyth, R., G. Jacobs, and H. Rogers (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society* 32, 329–350.
- Snidecor, J., C. (1951). The pitch and duration characteristics of superior female speakers during oral reading. *Journal of Speech and Hearing Disorders* 16, 44–52.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America* 85, 1699–1707.
- Trautmüller, H. and A. Eriksson (1995). The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America* 97, 1905–1915.
- Wolfe, V. I., D. L. Ratusnik, F. H. Smith, and G. Northrop (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders* 55, 43–50.
- Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, and R. Temple (Eds.), *Papers in Laboratory Phonology VI*, pp. 75–87. Cambridge: Cambridge University Press.



# Accent in L2 — An acoustic analysis of the vowels of German speakers of English

**Katrin Liebthal**

*Institut für Phonetik und digitale Sprachverarbeitung, Christian–Albrechts–Universität  
Kiel, German*

---

This study assessed the effect of English language experience on native German speakers' production of English vowels. The speech data was analysed from two speakers of German (an experienced and an inexperienced speaker) as well as from a control group of native English speakers recorded as part of the MAR-SEC database. The subjects' accuracy in producing the English vowels /i/, /ɪ/, /e/, /æ/, /ʌ/, /æ/ was assessed acoustically. Their production of German vowels was also compared to the production of their English vowels. The experienced German subject produced the English vowels more accurately i.e., in a more native manner, than the inexperienced German subject. The results show differences between the German and English speakers' vowels as well as for the category formation strategy between the German subjects. The experienced speaker uses one category for both languages, whereas the inexperienced subject seems to have established a new category for some English vowels.

---

## 1. Introduction

Are adults able to learn foreign speech sounds authentically? Can their success or failure be accounted for by corresponding sounds in their native language and the L2? The aim of many speech learning models is to answer these questions by establishing different hypotheses. The Critical Period Hypothesis (Lenneberg 1967) states that it is impossible for a learner to learn a foreign language to a native standard after the so called "critical period" has passed, i.e., when he or she is beyond a certain age. The particulars of this period range from 2 to 16 years. Other scientists do not believe in the existence of such a critical period. They assert that language acquisition depends upon and is influenced by other factors. The Speech Learning Model (SLM) proposed by James Flege (2003) is based on the relationship between L1 and L2 as well as how they can influence each other. The model predicts that adult L2 learners will produce new L2 sounds – that is, L2 sounds that don't occur in L1 – more authentically because extended L2 experience leads to the establishment of phonetic categories for L2 sounds that do not have a corresponding L1 sound. Similar sounds on the other hand will stay accented even after a longer exposure to the L2. This is due to a blocking of category formation for similar sounds through the so-called "equivalence classification", a mechanism which causes similar sounds in

the L1 and L2 to be treated as belonging to the same phonetic category (Flege and Bohn 1992).

The experiments reported here examine the influence of L2 experience on the production of L2 monophthongs that do not always have an acoustic counterpart in the L1. The hypothesis to be tested takes its pattern from experiments by Flege and Bohn (1992): whereas L2 experience does not affect the production of the similar English vowels /i, ɪ, ε/ by German speakers, an English /æ/, which has no German counterpart and can thus be classified as a “new” sound, will be produced more authentically by experienced German speakers of English than by inexperienced ones. To test this hypothesis, the production of the English vowels /i, ɪ, ε, æ/ as produced by two groups of differently experienced German speakers of English was compared to the production of these vowels by a native English control group. There was also a German control group to see if the English vowels produced by the German speakers were similar to Standard German vowels or would show different values. The questions that motivate this experiment are:

1. Can the phenomena of interlingual identification and equivalence classification be found in this analysis, i.e., do speakers replace similar sounds with the counterparts of their native language?
2. Are new sounds produced in a native manner?
3. What influence does language experience have? Do more experienced speakers achieve better, i.e., more authentic results than inexperienced speakers?

It is generally assumed that adults are not able to produce the sounds of an L2 authentically, i.e., like native speakers. On the other hand, Flege (1987a, 1987b, 1992) has shown that adults may succeed in that task if the L2 vowel is sufficiently different from the vowels of the L1, and if the L2 learner has been exposed for long enough to the L2. Flege (1987b, 1990, 1992) drew from these findings the conclusion that new vowels evade the process of equivalence classification and that enough input is sufficient to enable learners to build phonetic categories for new vowels. Flege and Bohn (1990) also found that intensive L2 contact not only improves L2 production but also the perception of vowel differences such as the contrast between similar and new vowels, e.g., English /ε/ vs /æ/. Adult learners are able to establish phonetic categories for new L2 sounds and thus produce the sounds authentically. Similar sounds, on the other hand, will be produced with an accent even after long and intensive exposure to the L2 as the building of a category is blocked by equivalence classification.

In the first part of the experiment, the acoustic qualities of the English vowels /i, ɪ, ε, æ, ɑ:, ɜ:/ as produced by native English speakers as well as by native German speakers were examined and compared.

In the second part of the experiment, the acoustic qualities of the English vowels /i, ɪ, ε, æ, ɑ:/ as produced by native German speakers and the acoustic qualities of the German vowels /i, ɪ, ε, ε:, ɑ:, a, e:/ as produced by the same native German speakers were examined and compared.

## **2. Method**

### *2.1 Subjects*

Speech production data were collected from three groups of adult subjects. Two groups of L1 German speakers with one female speaker each differed primarily in terms of English language experience. The subject in the “inexperienced” L1 German group (jli) had been exposed to English only at school, whereas the experienced German subject (baf) had, as well as having learnt English at school, spent one year living and working in the UK. Both German speakers were under the age of 30 and had been raised in Northern Germany. The sort of English taught at their schools was Southern Standard British English. The third group consisted of female English speakers from the MARSEC database (**MA**chine- **R**eadable **S**poken **E**nglish **C**orpus) (Deterding 1997).

### *2.2 Materials*

The German speakers produced both German and English materials. For the German materials, the subjects read 29 German sentences consisting of the frame sentence “Ich sage X.”, where X was a two-syllable German word with stress on the first syllable. (X was also nuclear accented). The vowels in the initial stressed syllable of X were one of the monophthongs /i/, /ɪ/, /ɛ/, /a/, /e:/, /ɛ://. The sentences were presented in randomized order. For vowels, number of tokens and words, see Table 1.

For the English data, the subjects read 30 English sentences consisting of the frame sentence “I will say X.”, where X was a two syllable English word with stress on the first syllable and X being nuclear accented as well. The vowels in the initial stressed syllable of X were one of the monophthongs /a:/, /i:/, /ɛ/, /æ/ und /ɪ/. The sentences were presented in randomized order. For vowels, number of tokens and words, see Table 2.

In addition to the English sentences, the subjects were asked to read an English text taken from the MARSEC database, consisting of prayers and a reading from the Bible. This text was chosen as it contained hardly any foreign words and was thus relatively easy to read for the German subjects. From these data, the formant values of the stressed English vowels /i, ɪ, ɛ, æ, aɪ, ʌ, ɜ:/ were taken from both subjects. For vowels, number of tokens and words, see Table 3.

The German and English sentences were formed in such a way that the target vowels were surrounded by a similar context in both languages. This was done to minimise coarticulation effects (see Table 4).

According to Flege and Bohn (1990), a review of the literature does not clearly indicate whether the /i/ and /ɪ/ vowels of English are perceptually similar or identical to

| Vowel | Number of tokens | Word                                      |
|-------|------------------|---|
| /ɑː/  | 5                | large, heart, father, part, start         |
| /iː/  | 6                | meet, keep, cheat, need, beat, heat       |
| /ɛ/   | 5                | bet, head, deck, leg, bed                 |
| /æ/   | 6                | bad, back, gag, cap, map, bat             |
| /ɪ/   | 8                | sin, kick, bit, pit, chin, bin, list, big |

Table 1: Vowels, number of tokens and words from the English word list.

| Vowel | Number of tokens | Word   |
|-------|------------------|--|
| /ɑː/  | 5                | Vase, Nase, Vater, Dame, Hase                      |
| /eː/  | 6 (5)            | beten, Läden, lesen, reden, (Räder)                |
| /ɛː/  | 1 (2)            | bäten, (Räder)                                     |
| /iː/  | 7                | bieten, Lieder, Liebe, liegen, Miete, Wiese, Krise |
| /a/   | 3                | backen, Matten, Ratten                             |
| /ɛ/   | 4                | Kästen, retten, meckern, betten                    |
| /ɪ/   | 4                | Kinder, bitten, kippen, Mitte                      |

Table 2: Vowels, number of tokens and words from the German word list.

German /i/ and /ɪ/. German and English /ɛ/ are probably identical, while English /æ/ can be regarded as a new sound for speakers of Standard German (Flege and Bohn 1990).

In both languages, the high front vowels /i/ and /ɪ/ differ in terms of spectral quality and duration (Flege and Bohn 1990). Stevens (1959a, 1959b) showed that for native English listeners, the /i/ - /ɪ/ contrast is cued primarily by spectral differences, while duration is a less important cue.

According to (Flege and Bohn 1990), some acoustic studies (Jones 1960; Moulton 1962) have supported the auditory impression that German /i/ and /ɪ/ are somewhat more peripheral than English /i/ and /ɪ/, but Disner's (1983) acoustic comparison of English and German /i/ revealed no significant differences. Delattres's (1964) observation that short vowels such as /ɪ/ are more central in English and more open in German than long vowels such as /i/ does not seem to be true of all German dialects (Flege and Bohn 1990).

Like English /i/ and /ɪ/, English /ɛ/ also has an identifiable counterpart in German (Flege and Bohn 1990). The /ɛ/ of both languages is short and nominally lax. According to (Flege and Bohn 1990), no information on acoustic differences between English and German /ɛ/ is available. He concludes that since English /ɛ/ is not mentioned as a problem in textbooks for German learners of English, English and German /ɛ/ may in fact be identical (Flege and Bohn 1990).

Concerning the vowel /æ/, Standard German has no long vowel in the low front portion of the vowel space (Flege and Bohn 1990). According to Flege and Bohn (1990), the

| Vowel | Number of tokens | Word  |
|-------|------------------|---|
| /æ/   | 15               | thank (2x), mankind, can, Saint, Matthew, chapter, man, gathered, hand (3x), compassion, understanding, planning  |
| /ɛ/   | 35               | heavens, flesh, everyone, twenty-five, said, separate, shepherd, separates, left (2x), blessed, inherit, welcomed (2x), brethren, devil, welcome, help (5x), footsteps, remembering, self-forgetting, helper, strengthen, repent, respect, forget, myself, improsperity, against, let, cherrish |
| /i:/  | 39               | dealings, feet, Jesus (3x), even, speaking, need (2x), reaches, reading, sheep (2x), me (7x), see (3x), thee (7x), least (2x), these (2x), people, weaknesses (2x), reach, keep   |
| /ɪ/   | 27               | beginning, image, written, mystery, living, kingdom (2x), sit, king, drink (2x), sick (2x), visited, prison (3x), visit, minister, humility, forgiveness, wisdom, sins, resist, Christian (2x), risks   |
| /ʌ/   | 23               | mother, lump, coming, nothing, son, comes, another, but, come (3x), hungry (3x), son's, love (3x), sufferings, above, touch, others, suffering  |
| /ɑ:/  | 15               | answer (2x), father (3x), depart, ask (4x), heart, past, guard, last  |
| /ɜ:/  | 15               | earth (2x), word (4x), church, world, thirsty (3x), cursed, eternal (2x), hurt  |

Table 3: Vowels, number of tokens and words from the English text.

| German  |        | English |        |
|---------|--------|---------|--------|
| Phoneme | Word   | Phoneme | Word   |
| /i:/    | mieten | /i:/    | meet   |
| /ɪ/     | bitten | /ɪ/     | bit    |
| /e:/    | beten  | /ɜ:/    | earth  |
| /ɛ:/    | bäten  | /æ/     | bad    |
| /ɛ/     | retten | /ɛ/     | read   |
| /ɑ:/    | Vater  | /ɑ:/    | father |
| /a/     | Matte  | /ʌ/     | mother |

Table 4: Phoneme - Word tabular of German and English words occurring in this experiment

entire inventory of German monophthongs consists of pairs governed by the opposition long-close vs. short-open, so that a long vowel like /æ/ that is low and front runs counter to the organizing principle of German vowels. For most speakers of German, long an front implies close (Flege and Bohn 1990). In a listening test, Flege and Bohn (1990) found out that German subjects identified English /æ/ in terms of German /ɛ:/.

English does not use the opposition lax-tense (realized phonetically as short-and-close vs. long-and-open) to distinguish two different low centered vowel phonemes such as /a/ and /a:/ (Moulton 1962).

In both languages, there are phonological oppositions such as front vs. back, open vs. closed, rounded vs. unrounded, tense vs. lax and long vs. short.

By comparing these oppositions between the two languages, it can be found that German decentralized vowels /i, e, a, o, u, y, ø/ are tense, whereas the centralized vowels /ɪ, ɛ, ə, ɔ, ʊ, ʏ, œ/ are lax. English does not have a continuous opposition between tense and lax vowels, but only sometimes uses muscular contraction to differentiate between pairs such as /i/ vs. /ɪ/, /e/ vs. /ɛ/, /u/ vs. /ʊ/ and /o/ vs. /ɔ/ (Kufner 1971). So German /i/ is generally tenser than its English equivalent, whereas English /i/ is more open and shorter than German /i/ (Moulton 1962).

English does not use the opposition lax-tense (realized phonetically as short-and-close vs. long-and-open) to distinguish two different low centered vowel phonemes such as /a/ and /a:/ (Moulton 1962).

English short vowels /e/ and /æ/ are often not distinguished by German speakers but produced as a half-open vowel between /e/ and /æ/ that is qualitatively similar to German half-open short /ɛ/ (Weiher 1975). So concerning the opposition of front vs. back, English /æ/ is fronter than German /a/.

### 2.3 Data processing and labelling

The speech data processing programme *EMU* (Cassidy 1998; Cassidy and Harrington 2001) was used for the processing of the data. The German and English data from the German speakers were available as waveforms and spectrograms.

The relevant accented vowels /i, i:, ɪ, ɛ, æ, a:, ɑ:, ʌ, ɜ:/ were labelled manually, the first four formants were calculated automatically using an LPC analysis program (LPC order of 16, pre-emphasis -0.95, a Hamming window of 20 ms and a shift of 5 ms). Problems in the automatic formant tracking were corrected by hand.

For the measurement of the first two formants, a label was set in the temporal middle of each relevant monophthong. The formant values were then extracted at this point using LPC analysis.

The MARSEC data had already been processed in the same way. For the acoustic analysis as well as the statistical tests, the *EMU* speech database analysis program (Cassidy and Harrington 1996) was used.

### 3. Results

ANOVAs were carried out on the data. In all tests, the dependant variables were the formant values of the first and second formant of the respective vowel. The independent variables were language (German or English native language), type of text (text or word list) and speaker group (MARSEC, baf or jli). The ANOVAs were carried out for each type of vowel separately.

If the German speakers produce the English sounds in a native manner, the measurements will not differ significantly between the German and English speakers. If the results show significant differences between the two languages, the English sounds of the German speakers will not be produced in a native manner. Therefore, the two different speaker groups, i.e., English (MARSEC) and German native speakers (regardless of proficiency), are compared with each other. The independent variable is the different native language of the speakers, i.e., English and German. The two factor ANOVA revealed that the vowels /ɑ:/ and /a:/ as well as /ʌ/ and /a/ are not significantly different between the three speaker groups. They can be regarded as being similar. The two German speakers achieve qualities that are statistically equivalent to native.

If the English vowels of the German speakers do not differ significantly from their German vowels, the German speakers use their German vowel quality to produce an English vowel. To test this hypothesis, the English data of the German speakers are compared with their German data. The relevant variable is the English or German language used by the German speakers. The difference between F1 and F2 values for the vowels /æ/ (F1:  $F = 195, p < 0.001$ ; F2:  $F = 21.1, p < 0.001$ ), /ɛ/ (F1:  $F = 18.65, p < 0.001$ ; F2:  $F = 28.7, p < 0.001$ ) and /ɜ:/ (F1:  $F = 13.2, p < 0.001$ ; F2:  $F = 29.8, 0.001$ ) is highly significant between the two languages. At least /æ/ and /ɜ:/ are seen by Flege and Bohn (1992) and Flege, Schirru, and MacKay (2003) as “new” phones and should thus be produced in a native manner by the experienced German speaker. Since the differences between these vowels are highly significant, they offer no support for this hypothesis. Differences in the vowel /i:/ are highly significant only between the speakers and only for F2 ( $F = 32.3, p < 0.001$ ), but not between the languages, which suggests that this difference is speaker- but not language-specific. Flege and Bohn’s (1992) finding – that the German vowel /i:/ is spectrally similar but not the same as the English vowel /i:/ – seems to be confirmed here.

The last hypothesis to be tested was that the experienced speaker would produce the “more authentic” English vowels, especially in the case of the vowels /æ/ and /ɜ:/. Therefore, the German speakers are compared with each other. The relevant variable is the degree of experience of the German speakers. The inexperienced German speaker made a significant distinction between the vowels /ɑ:/ and /a:/ in the German and the English word lists (F2:  $F = 0.0007, p < 0.001$ ). She must have used or established different phonetic categories for these vowels. The experienced German speaker seems to possess or to use only one category for these vowels as shown as she shows no differences between the German and English production of either /i:/ or /a:/. ish production of either /i:/ or /a:/. However, her productions of the English vowels /ɪ/ (F1:  $F = 2.1e - 14, p < 0.001$ ; F2:

0.0014,  $p < 0.01$ ), /ʌ/ (F1:  $F = 0.009$ ,  $p < 0.01$ ), /ɛ/ (F1:  $F = 0.0002$ ,  $p < 0.01$ ; F2:  $F = 1.7e - 07$ ,  $p < 0.001$ ), /æ/ (F1:  $F = 8.9e - 16$ ,  $p < 0.001$ ; F2: 0.0002,  $p < 0.001$ ) are significantly different from their German counterparts. The inexperienced speaker produces all English vowels significantly different from her German vowels.

#### 4. Analysis

The hypothesis underlying the analysis was as follows: if the German speakers produce the English sounds in a native manner, the measurements will not differ significantly between the German and English speakers. If the results show significant differences between the two languages, the English sounds of the German speakers will not be produced in a native manner.

If these hypotheses are supported, it might be the case that the German speakers pronounce their English vowels in a German manner, i.e., should their English vowels not differ significantly from their German vowels, the German speakers use their German vowel quality to produce an English vowel.

Should there also be differences in the pronunciation of their English vowels between the two German speakers, it is hypothesized that the experienced speaker would produce the “more authentic” English vowels, especially in the case of the vowels /æ/ and /ɜ:/.

The analysis was divided into three parts:

1. The two different speaker groups, i.e., English (MARSEC) and German native speakers (regardless of proficiency), are compared with each other. The independent variable is the different native language of the speakers, i.e., English and German.
2. The German speakers are compared with each other. The relevant variable is the degree of experience of the German speakers.
3. The English data of the German speakers are compared with their German data. The relevant variable is the English or German language used by the German speakers.

ANOVAs were carried out on the data to examine the questions above. In all tests, the dependent variables were the formant values of the first and second formant of the respective vowels. The independent variables were language (German or English native language), type of text (text or word list) and speaker group (MARSEC, baf or jli). The ANOVAs were carried out for each vowel type separately.

For the first analysis, the speakers were divided into two groups: the MARSEC-speakers and the native German speakers. It was analysed how the different native languages of the two speaker-groups affect the production of the English vowels.

Figure 1 shows the Distribution of the F1- and F2-values of the English vowels of the MARSEC-speakers and both native German speakers. Even without a statistical analysis, the diagramme shows that the vowels /ɛ/ and /æ/ of the German speakers (displayed as letters E and A in the diagramme) have overlapping ellipses, whereas these vowels produced by the native English speakers are clearly separated from each other. This overlap suggests that the German speakers don't differentiate distinctly between the two vowels in their production.

The ANOVA for these vowels confirms this interpretation: for the vowel /æ/, there is a highly significant difference for F1 as well as for F2 between the languages ( $F = 195, p < 0.001$ ;  $F = 21.1, p < 0.001$ ). This is also true for the vowel /ɛ/ ( $F = 18.65, p < 0.001$ ;  $F = 28.7, p < 0.001$ ). The vowel /i/ shows no significant difference between the languages but a highly significant difference for F2 between the speakers ( $32.3, p < 0.001$ ). This means that the two German speakers differ in their production of this vowel with each other but that there is no difference between the German and the English speakers. A highly significant difference between the languages can be found for the vowel /ɪ/ for F1 as well as for F2 ( $F = 89.6, p < 0.001$ ;  $F = 15.2, p < 0.001$ ). The vowels /ɑ:/ and /ʌ/ show no significant difference whatsoever which means that there are produced "native-like" by the German speakers. The ANOVA of the vowel /ɜ:/ shows a highly significant difference for the languages for F1 and F2 ( $F = 13.2, p < 0.001$ ;  $F = 29.8, p < 0.001$ ).

The second part of the analysis looks at the vowel production the two German speakers, especially the influence of text kind and speaker on the production of the English vowels. Differences are made between the experienced and the inexperienced speaker as well as between their production of the vowels in the English wordlist and in the English text.

Figure 2 shows the formant ellipses of the English vowels from the text and the wordlist, produced by the two German Englishspeakers. At first sight, the two speakers seem to produce the vowels quite similarly, which proves wrong. The statistical analysis reveals a significant difference between the two text kinds (i.e. wordlist and text) for the vowel /æ/ for F1 and F2 ( $F = 6.1, p < 0.05$ ;  $F = 5.6, p < 0.05$ ). For the same vowel, there is a highly significant difference for F1 between the speakers ( $F = 13, p < 0.001$ ) as well as for F1 for the interaction between text kind and speakers ( $F = 12.8, p < 0.001$ ). This means that both speakers make a distinct difference in their production of the vowel /æ/ between the wordlist and the text. Furthermore, they produce this vowel differently from each other. The vowel /i/ differs significantly in F2 between the two text kinds ( $F = 4.9, p < 0.05$ ) and highly significant in F2 between the two speakers ( $F = 17.6, p < 0.001$ ). Significant is also the interaction between text kind and speaker for F2 ( $F = 4.6, p < 0.05$ ).

The results of the ANOVAs can be summarized as follows:

The *Welch Two Sample t-test* revealed that the LMS speaker had significantly greater distances to the midpoint than the MMS speaker ( $t = 2.5958, df = 29, 12, p < 0.05$ ).

- The two factor ANOVA revealed that the vowels /ɑ:/ and /a:/ as well as /ʌ/ and /a/ are not significantly different between the three speaker groups (/ɑ/:  $F = 1, p >$

0.05; /ʌ/:  $F = 0.7, p > 0.05$ ). They can be regarded as being similar. The two German speakers achieve qualities that are statistically equivalent to native.

- The inexperienced German speaker made a significant distinction between the vowels /ɑ:/ and /a:/ in the German and the English word lists (F2:  $F = 0.0007, p < 0.001$ ). She must have used or established different phonetic categories for these vowels. The experienced German speaker seems to possess or to use only one category for these vowels as shown by ...???
- The difference between F1 and F2 values for the vowels /æ/ (F1:  $F = 195, p < 0.001$ ; F2:  $F = 21.1, p < 0.001$ ), /ɛ/ (F1:  $F = 18.65, p < 0.001$ ; F2:  $F = 28.7, p < 0.001$ ) and /ɜ:/ (F1:  $F = 13.2, p < 0.001$ ; F2:  $F = 29.8, 0.001$ ) is highly significant between the two languages. At least /æ/ and /ɜ:/ are seen by Flege and Bohn (1992) and Flege, Schirru, and MacKay (2003) as “new” phones and should thus be produced in a native manner by the experienced German speaker. Since the differences between these vowels are highly significant, they offer no support for this hypothesis.
- Differences in the vowel /i:/ are highly significant only between the speakers and only for F2 ( $F = 32.3, p < 0.001$ ), but not between the languages, which suggests that this difference is speaker- but not language-specific. Flege and Bohn’s (1992) finding — that the German vowel /i:/ is spectrally similar but not the same as the English vowel /i:/ — seems to be confirmed here.
- The experienced speaker shows no differences between the German and English production of either /i:/ or /a:/. However, her productions of the English vowels /ɪ/ (F1:  $F = 2.1e - 14, p < 0.001$ ; F2:  $0.0014, p < 0.01$ ), /ʌ/ (F1:  $F = 0.009, p < 0.01$ ), /ɛ/ (F1:  $F = 0.0002, p < 0.01$ ; F2:  $F = 1.7e - 07, p < 0.001$ ), /æ/ (F1:  $F = 8.9e - 16, p < 0.001$ ; F2:  $0.0002, p < 0.001$ ) are significantly different from their German counterparts.
- The inexperienced speaker produces all English vowels significantly differently from her German vowels.

## 5. Discussion

The experiments reported here were based on studies by Flege and Bohn (1990, 1992), Flege, Munro, and MacKay (1996) as well as Flege, Bohn, and Jang (1997), but the pattern of results is very different from those in their experiments.

According to Flege and Bohn (1990, 1992), the experienced speaker should have produced the English vowels /æ/ and /ɜ:/, classified by Flege as “new” for German native

*Accent in L2*

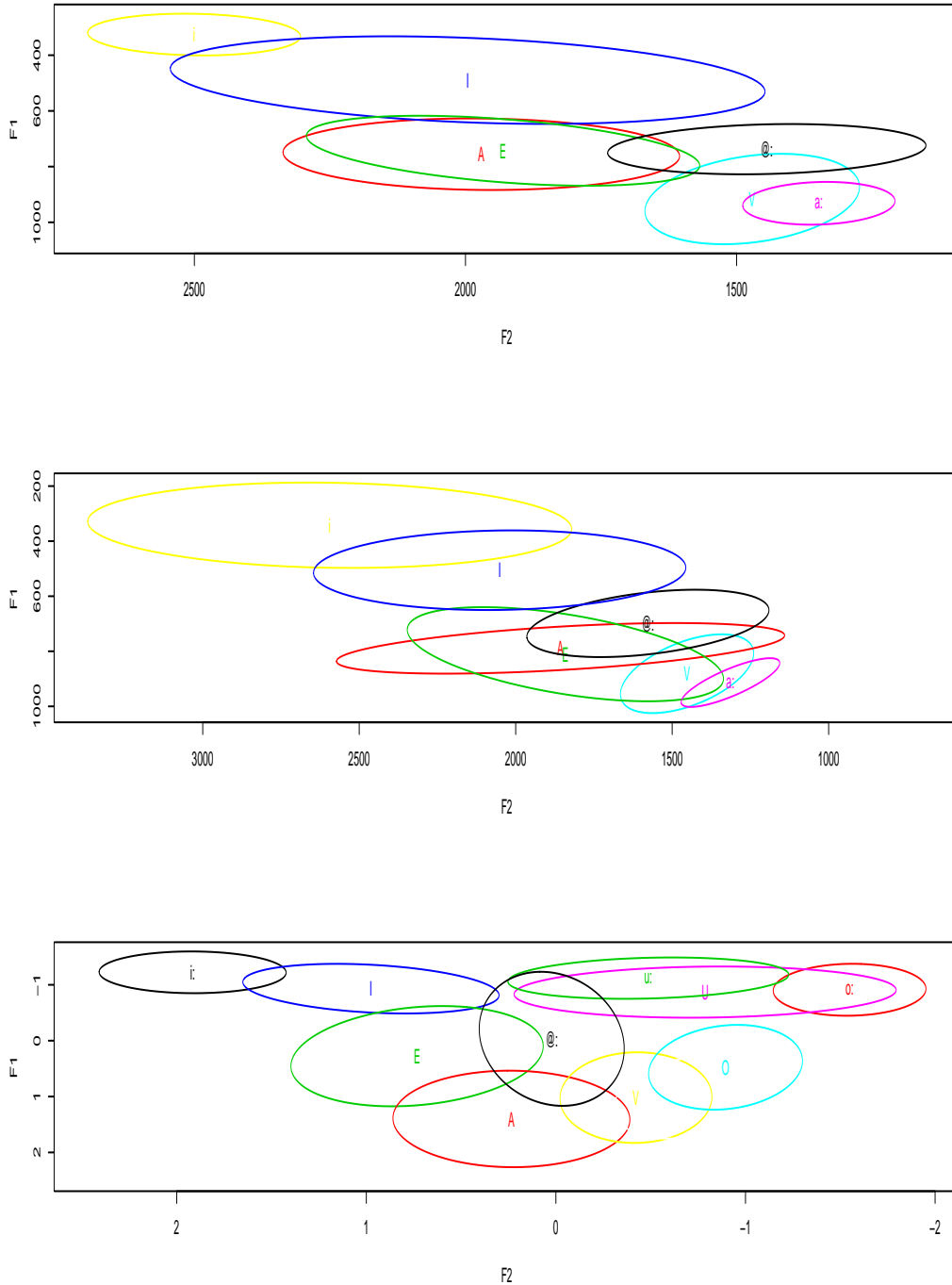


Figure 1: Distribution of the F1- and F2-values (Hz) of the English vowels from the English text of the experienced speaker (above), the inexperienced speaker (middle) and the MARSEC-speakers (below)

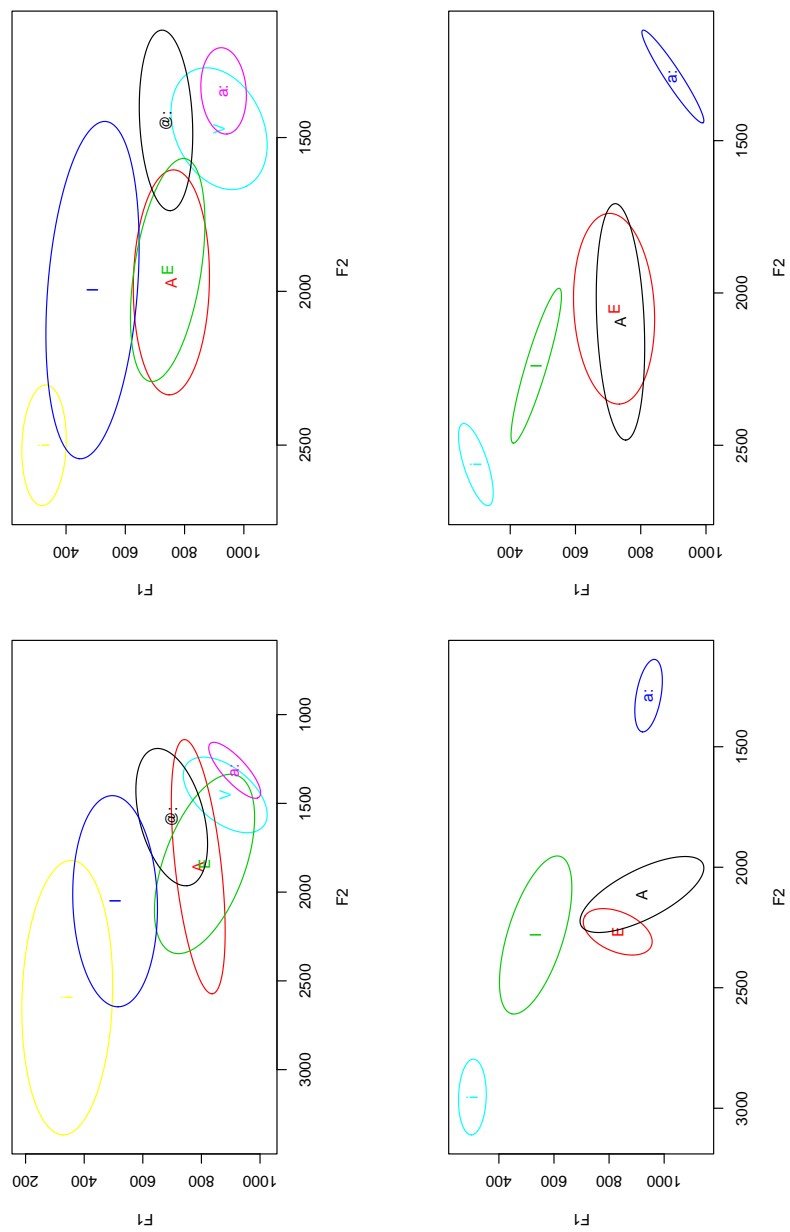


Figure 2: Distribution of the F1- and F2-values (Hz) from the English text (above) and the English wordlist (below) of the experienced speaker (left) and the inexperienced speaker (right)

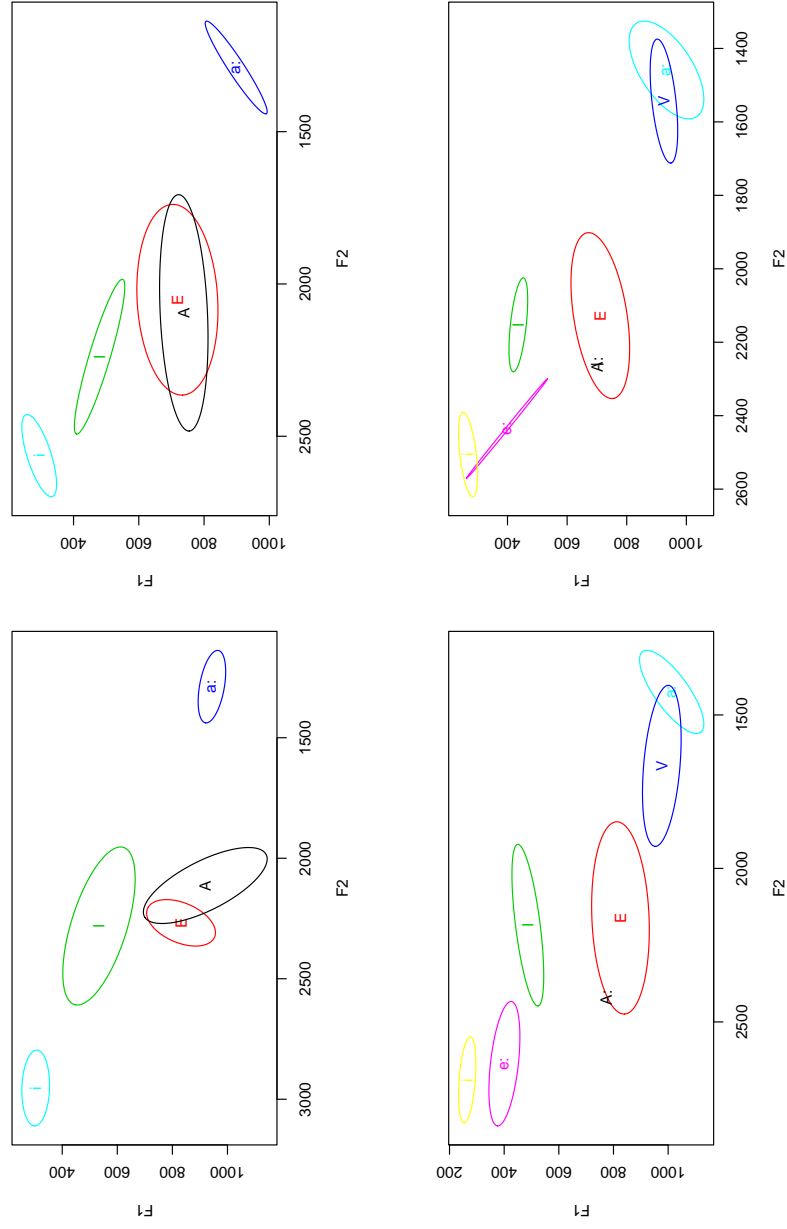


Figure 3: Distribution of the F1- and F2-values (Hz) from the English wordlist (above) and the German wordlist (below) of the experienced speaker (left) and the inexperienced speaker (right)

speakers, better i.e., in a more native manner than the inexperienced speaker. But this was not found to be the case in this study.

Results from earlier studies (Weiher 1975; Flege and Bohn 1992) have already established that the production of foreign, i.e., “new” sounds in the L2, especially the target-like production of the vowel /æ/, poses a considerable problem for German native speakers.

The vowel /ɛ/, classified by Flege and Bohn (1990, 1992) as “similar”, and assumed to cause no problems for German native speakers, also showed highly significant differences between the two languages. The German speakers failed to produce the English sound in a native manner.

The results for the vowels /ɑ:/ and /a:/ from this experiment are interesting, however: whereas the experienced speaker uses the same vowel in both languages (and which is not differentiated from the English speakers’ /ɑ:/), the inexperienced speaker seems to have established a new category for this vowel. Her realizations of the English vowel do not differ significantly from the English speakers’ but from the German counterpart. Unfortunately, there is a lack of information about how this vowel differs phonetically in English and German and so it is difficult to explain this result further.

There are several possible interpretations for the differences between these findings and those from Flege’s studies (1990, 1992, 2003). One possibility is that the experienced speaker had not mastered the English language well enough. Whereas the experienced speakers in Flege’s experiments had spent several years living abroad, she had only lived there for one year, this year furthermore predating the experiment by two years. Though she had spoken English a great deal in the interim, her use of German was greater, leaving room for Flege’s theory of the influence of the L2 through the mother tongue (Flege 1987a). The most important criteria for achieving an almost native-like production seem to be the length of the stay abroad and direct contact with the foreign language. In Flege’s studies, the inexperienced speakers had learnt English at school for about six years and had spent half a year in an English-speaking country. In the experiment reported here, it was the experienced speaker who met these criteria. The experienced speakers in Flege’s experiments, however, had been living in an English-speaking environment for at least six years. These differences indicate that adults need quite a long time to show signs that they have acquired a new L2 vowel phonetically (Flege and Bohn 1992). Even then, regularities of the L1 can influence the production of the new vowel.

As no discrimination experiment was carried out, it is impossible to say whether the experienced speaker was able to differentiate the relevant English sounds, especially /æ/ and /ɛ/. There was also only one speaker in the present study, whereas the number of speakers in Flege’s experiments was much greater. Due to the small amount of data in this experiment, the results cannot be called representative. A further perception experiment could have been carried out for this study, letting native speakers of English rate the data of the German speakers of English. Only one native English speaker listened to the data before the statistical analysis. He made a general statement that the experienced speaker had a more marked English accent than the inexperienced speaker.

## 6. Conclusion

The experiments presented here show some results that might be of interest for further research. To make a more general statement about accent in the L2, a larger group of native German speakers should be recorded than just two people. Furthermore, there should be at least two listening experiments which examine the ability of the speakers to distinguish between different sounds of the L2 and which also have the data of the German speakers of English rated by native English speakers. But even if Flege's studies (1987b) show that L2 learners produce "new" sounds in a way that they are rated as being authentic by native speakers, it is extremely difficult to examine whether or not a correct, new phonetic category has been established (Tsukada 1999). According to Tsukada (1999), another question still remains unanswered: is category formation a prerequisite for an authentic, accent-free L2 production? Considering the results for the vowel /a:/ for both German speakers in this experiment, the answer to this question — though cautiously formulated — might be, "No". In the production of this vowel, neither speakers differed significantly from the native English speakers. But only the inexperienced speaker seems to have established a new category, whereas the experienced speaker uses one category for both languages. There seem to be speakers who only need one category for certain vowels and who nevertheless are able to produce these vowels without an accent in the L2, whereas other speakers use two categories for these vowels. It could be of future interest to examine these findings further and to find out the extent to which this phenomenon really is speaker- or perhaps vowel-dependent.

## References

- Cassidy, S. (1998). <http://www.shlrc.mq.edu.au/emu>.
- Cassidy, S. and J. Harrington (1996). EMU: An enhanced hierarchical speech data management. In P. McCormack and A. Russel (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*, pp. 361–366. Adelaide: ASSTA.
- Cassidy, S. and J. Harrington (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication* 33, 61–77.
- Delattre, P. (1964). Comparing the vocalic features of English, German, Spanish and French. *International Review of Applied Linguistics* 2, 71–97.
- Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association* 27(1), 47–55.
- Disner, S. F. (1983). Vowel quality: The relationship between universal and language specific factors. *University of California at Los Angeles Working Papers in Phonetics* 58.
- Flege, J. E. (1987a). A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8(2), 162–177.
- Flege, J. E. (1987b). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

- Flege, J. E. and O.-S. Bohn (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11, 303–328.
- Flege, J. E. and O.-S. Bohn (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14, 131–158.
- Flege, J. E., O.-S. Bohn, and S. Jang (1997). Effects of experience on non-native speaker's production and perception of English vowels. *Journal of Phonetics* 25, 437–470.
- Flege, J. E., M. J. Munro, and I. R. A. MacKay (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17, 313–334.
- Flege, J. E., C. Schirru, and I. R. A. MacKay (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication* 40(4), 467–491.
- Jones, D. (1960). *An outline of English phonetics* (9th ed.). London: Heffer.
- Kufner, H. L. (1971). *Kontrastive Phonologie Deutsch-Englisch*. Stuttgart: Ernst Klett Verlag.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Moulton, W. G. (1962). *The Sounds of English and German. A Systematic Analysis of the Contrast between the Sound Systems*. Chicago: The University of Chicago Press.
- Stevens, K. (1959a). Effect of duration upon vowel identification. *Journal of the Acoustical Society of America* 31, 109.
- Stevens, K. (1959b). The role of duration in vowel identification. *M.I.T. Research Laboratory of Electronics, Quarterly Progress Report* 52, 136–139.
- Tsukada, K. (1999). *An Acoustic Phonetic Analysis of Japanese-accented English*. Ph. D. thesis, Speech, Hearing and Language Research Centre (SHLRC), Department of Linguistics, Macquarie University, Sydney, Australia.
- Weiher, E. (1975). *Lautwahrnehmung und Lautproduktion im Englischunterricht für Deutsche*. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel Nr.3.

# An EMMA and EPG study on token-to-token variability

**Christine Mooshammer**

*Institut für Phonetik und digitale Sprachverarbeitung, Christian-Albrechts-Universität Kiel, Germany*

**Pascal Perrier**

*Institut de la Communication Parlée, UMR CNRS 5009, INPG & Université Stendhal Grenoble, France*

**Susanne Fuchs, Christian Geng and Daniel Pape**

*ZAS – Research Centre for General Linguistics, Berlin, Germany*

---

The potential role of physical correlates of articulatory variability in vowel production was studied by means of EMMA and EPG. In accordance with Perkell & Nelson (1985), the hypothesis to be tested is that the amount and shape of token-to-token variability is influenced by the amount of palatal contact in such a way that high vowels exhibit different variability patterns than the other vowels. 14 German vowels of three speakers of German were recorded with and without bite-block and in bilabial and velar consonantal contexts repeated 10 times. Statistical results indicated that the amount of tongue variability, measured as the size of the dispersion ellipses describing tongue sensor positions, was constrained by the amount of palatal contact: variability was less when the amount of contacts was large. However, the three subjects exhibited different patterns of variability and only one of them showed a main orientation of the dispersion ellipses that was clearly different for high vowels in a sense that is compatible with data collected on speakers of American English by Perkell & Nelson. These results are discussed with regards to differences in the density of each language specific vowel inventory and the inter-individual variation among German subjects is furthermore attributed, at least partly, to morphological differences in the palatal shape and compensatory strategies.

---

## 1. Introduction

A classical and much debated approach in studying representations and motor control in speech production consists in looking for acoustic and/or articulatory invariants of phonemes (Stevens, 1972; Stevens & Blumstein, 1978; Blumstein & Stevens, 1979, 1980; Fujimura, 1986; Browman & Goldstein, 1985, 1986, 1990). However, the well-known noticeable variability of physiological, articulatory and acoustic signals of speech associated with coarticulation phenomena and variations in speaking style has led to an alternative approach that aims to study the variability itself. This approach consists in measuring the ranges of variation of the data (Wood, 1979; Perkell & Nelson, 1985; Folkins & Brown, 1987; Perkell, 1990; Beckman *et al.*, 1995), in looking at the internal structure of the data distributions

within the domain of variation (Kuhl, 1991, 1992), and in studying the possible relations between the density of phonological systems and the amplitude of the variability (Lubker & Gay, 1982).

Thus, physical correlates of phonemes are not considered to be strictly invariant but rather as regions of the motor, articulatory, acoustic, and/or perceptual spaces. Allophones of the same phoneme have to be located within these regions to allow communication between speakers and listeners. Keating (1990) used this concept in her window model to explain coarticulation in speech production. Guenther (1995), Guenther *et al.* (1998, 1999), Perkell *et al.* (1997) and Perrier (2003) also adopted the concept of regions to elaborate models of planning in speech production.

More generally, in order to study the interaction between perception and production in speech the characterization of the physical correlates of a phoneme in terms of range and nature of variation, rather than in terms of invariance, has been shown to be very fruitful. For instance, Lindblom (1990) within the context of the H & H and adaptive perception theories, suggested for vowels that formant patterns would be allowed to dramatically vary with speaking style (Hyper- versus Hypoarticulation), without endangering the perception, as long as the relative location of vowels in the overall vowel system are preserved. The underlying hypothesis is that the listener would adapt the size of her/his vowel space in the acoustic domain depending on the speaking style. Thus, different regions of the formants space would be associated with a same phoneme. This allows some freedom in the physical specification of the speech task, and, according to Lindblom (1990), this freedom would be used in the planning of speech sequences to minimize the articulatory effort while ensuring the perception of the articulated sound. The so-called perceptual magnet effect, introduced by Kuhl (1991, 1992) is also relevant for understanding how the negotiation between speaker-oriented criteria and listener-oriented requirements can operate. Studying the perception of phonetic categories, Kuhl provided evidence for an uneven structure of the perceptual region associated with each category: the closer a sound is to the perceptual prototype of a category the less it can be perceptually discriminated from this prototype. The proposed decrease of the perceptual discrimination ability away from the perceptual prototypes is very useful for explaining some aspects of the compensation strategies observed in speech production (Perkell *et al.*, 2000).

Thus, from the speech motor control perspective, Lindblom's and Kuhl's theories, both purely based on analyses of acoustic variability in relation to perception, help to understand the objectives of a speaker: they propose a number of constraints that speakers have to deal with, in order to ensure the perception of their speech, and they shed light on the freedom that speakers can use in order elaborate the planning of speech sequences. Consequently, an analysis of both, articulatory and acoustic variability allows us to find out more about the gestural accuracy required in order to match the perceptual constraints. From this perspective, Perkell & Nelson (1985) (see also Perkell, 1990) proposed a study that has formed the basis of many subsequent investigations.

Perkell & Nelson (1985) analyzed X-ray microbeam data in multiple repetitions of the vowels [i] and [a] in a variety of phonetic environments produced by three speakers of American English. For both vowels they found that the major axis of the dispersion ellipses characterizing the distributions of the pellets located in or close to the constriction region was parallel to the outline of the vocal tract walls. They explained these observations with the existence of passive "saturation effects" for these vowels, which would introduce strong non-linearities in the relations between muscle commands and articulatory position. In the case of [i], Perkell & Nelson (1985) suggested that the fact that the sides of the tongue blade are being pushed against the hard palate would strongly restrain tongue position variability in the direction perpendicular to the palate. According to these authors, in the case of [a], the

limitation of the variability perpendicularly to the pharyngeal walls would be due to a saturation in the length-tension characteristics of the active muscles. Perkell & Nelson (1985) interpreted their results in support of the Quantal Theory of speech production (Stevens, 1972), which suggests that the most frequent vowels in the world's languages inventory (and among them vowels [i] and [a]), would be articulated in regions of the vocal tract where articulatory changes would generate no or little auditory changes. In the case of [i] and [a], the "saturation effects" would limit the change in constriction area and then contribute to the perceptual stability of the articulated sound.

A similar experiment was carried out by Perkell & Cohen (1989) with a midsagittal electromagnetometer for vowels /i/, /a/ and /u/ in [bV<sub>1</sub>' V<sub>2</sub>b<sub>Λ</sub>] and [bV<sub>1</sub>'CV<sub>2</sub>b<sub>Λ</sub>] sequences, where C was one of the consonants [b, ʔ, h] and in which the influence of syllable stress was also considered. Only one point located on the dorsal tongue surface near the place of maximum constriction for /u/ was recorded and analyzed. The measured variability of this point did not exactly replicate Perkell and Nelson's (1985) findings, especially for [i]. However, given the fact that the points measured in Perkell & Nelson (1985) were more accurately located in the respective constriction location of each vowel, it can be considered that both experiments are consistent with each other and support the hypothesis of saturation effects limiting the variability in the direction perpendicular to the vocal tract midline. However, Perkell & Cohen's (1989) data show also that the phonetic context seems to influence the orientation of the dispersion ellipses, which tends to be "*rotated somewhat in the direction of the context vowel target location*" (Perkell, 1990, p.283), and to consistently vary with the intervocalic consonant.

Perkell & Nelson's (1985) experimental findings were replicated with different data sets and with a more systematic statistical approach by Beckman *et al.* (1995), and not only for /i/, /a/ and /u/, but for all vowels of English, including vowels that are not considered as quantal in the Quantal Theory.

In order to assess more precisely the respective contributions of contextual and token-to-token effects in the measured vowel variability, Hoole and Kühnert (1995) compared both effects for stressed vowels in German. They found that (a) consonantal context affects the target position of lax vowels to a greater degree than the one of tense vowels, (b) for lax vowels the amount of token-to-token variability was not influenced by vowel height or frontness and (c) tense front high vowels tended to vary to a lesser degree than tense back and low vowels. Considering Beckman *et al.*'s (1995) and Hoole and Kühnert's (1995) results, it can be argued that the patterns of articulatory variability measured for vowels could be largely determined by general mechanisms valid for every tongue and jaw gesture rather than by vowel specific saturation effects that would be used to satisfy specific perceptual constraints.

Possible biomechanical contributions to articulatory variability have been investigated by Shiller *et al.* (2002) who analyzed jaw positioning in vowel production. In a first experiment they measured token-to-token jaw variability in CVC sequences using an Optotrack system. The vowels /i/, /a/, /e/ and /æ/ in /k/, /t/, /s/ environment were recorded. In a second experiment, a computer-controlled robotic device coupled with the jaw was used to deliver mechanical perturbations to this articulator and estimate its stiffness in the mid-sagittal plane. They showed that the patterns of variability observed for the 4 vowels were consistent with the stiffness patterns. Indeed for each vowel, the variability was low in directions where stiffness was high and vice versa. This finding suggests that the contribution of mechanical properties of the articulators to the orientation of articulatory variability could be the main determining factor.

In this framework, the amount of token-to-token variability at target position could be seen as a consequence of the corruption of motor control signals by neural noise, from one

repetition of the task to the next. Harris and colleagues (Harris, 1998; Harris & Wolpert, 1998) have recently demonstrated the potential impact of such a noise on the end-position of target directed movements. They have proposed that motor control strategies underlying the production of such movements would even be organized in the aim to ensure the minimization of the thus induced variability at end position.

The study<sup>1</sup> presented in this paper is part of a larger project inspired by the objective to evaluate the possible influence of biomechanical factors on speech variability as compared to perception oriented factors. The token-to-token variability in the production of German vowels is measured in the articulatory and the acoustical domains. Based on Perkell & Nelson's (1985) conclusions, the following results were expected:

- Articulatory variability should decrease when the amount of palatal contacts increases.
- For high vowels, which have a high amount of palatal contacts, the major axis of the dispersion ellipses should be oriented along the vocal-tract midline.
- The articulation of quantal vowels should vary in constriction location (along the direction parallel to the vocal tract mid-line), but should be more constrained in constriction degree (along the direction perpendicular to the vocal tract mid-line).

Discrepancies between these predictions and our experimental results will be pointed out and they will be interpreted in terms of speech production control and/or of palatal influence on tongue positioning.

## 2. Method

### 2.1. Data Acquisition

Tongue, jaw and lower lip movements of three male speakers of Standard German were recorded by means of Electromagnetic Midsagittal Articulography (EMMA, AG100, Carstens). Tongue-palate contacts were recorded simultaneously with the Reading EPG system. Four sensors were attached to the tongue, one as far back as possible (TBACK), one close to the posterior border of the artificial EPG palate (TDORS), one approximately 1 cm behind the tongue tip (TTIP) and one equidistant between TDORS and TTIP (TBLADE). Two sensors on the nasion and on the upper incisors served as references for compensation of head movements relative to the helmet and definition of an intermediate coordinate system. Additional sensors were glued on the lower incisors, one on the vermillion border of the lower lip and another on the tongue tip, but they will not be considered here. The final coordinate system was defined by recordings of two sensors on a T-bar, manufactured individually for each subject in order to determine his bite plane. Original sample frequencies were 100 Hz for EPG data, 400 Hz for EMMA data and 48 kHz for the acoustical signal. For the analysis, the EMMA signals were low-pass filtered and downsampled to 200 Hz while the acoustical signal was resampled at 16kHz.

All subjects were recorded twice, once with a 5 mm thick bite block maintained between the second molars (hereafter BB condition) and once without bite block (henceforth, the normal condition). The BB condition was recorded in order to remove the contribution of the jaw to the token-to-token variability and to focus more specifically on the tongue control itself.

The material consisted of CVC<sub>3</sub> nonsense words with either velar or bilabial stops as consonantal context and one of the 14 German vowels /i:, ɪ, y:, ʏ, e:, ε, ø:, œ, a:, a, o:, ɔ, u:, u/. The initial stop was voiced and the medial voiceless. Examples of the target words

---

<sup>1</sup> The present paper is an extension of the paper presented in the *4th International Speech Motor Conference* held in June 2001 in Nijmegen, The Netherlands (Mooshammer et al., 2001)

are *gieke, gucke, büpe*. All nonsense words were embedded in the carrier sentence “Sage .... bitte” (“Say .... please”) and repeated 10 to 11 times.

## 2.2. *Measurements*

Vowel specific tongue positioning was determined visually using the following criteria: (a) the selected time-point had to be within the voiced part of the vowel and close to the acoustic vowel midpoint; (b) most of the tongue trajectories showed a turning point, which was then selected; (c) in the absence of a turning point, the time location of the minimum amount of EPG contacts was chosen. For the recordings without bite-block, the highest amount of jaw opening was also taken into account. Tongue sensor positions, EPG contact patterns and frequencies of the first and second formant were measured at this time instant.

To assess the token-to-token variability, two-sigma dispersion ellipses were computed for the 10–11 repetitions of each item. This gives four ellipses per vowel and speaker, i.e. 2 bite-block conditions and two consonant contexts. The ellipses were displayed in the sagittal plane for the three most posterior tongue sensors. Tongue tip sensor positions were not taken into account here because this articulator is not assumed to have a major influence on the production of vowels. The ellipses describe Gaussian estimations of the sensor positions distributions at the vowel target. Variability was measured on the basis of the area of these ellipses and of the angle of their major axes.

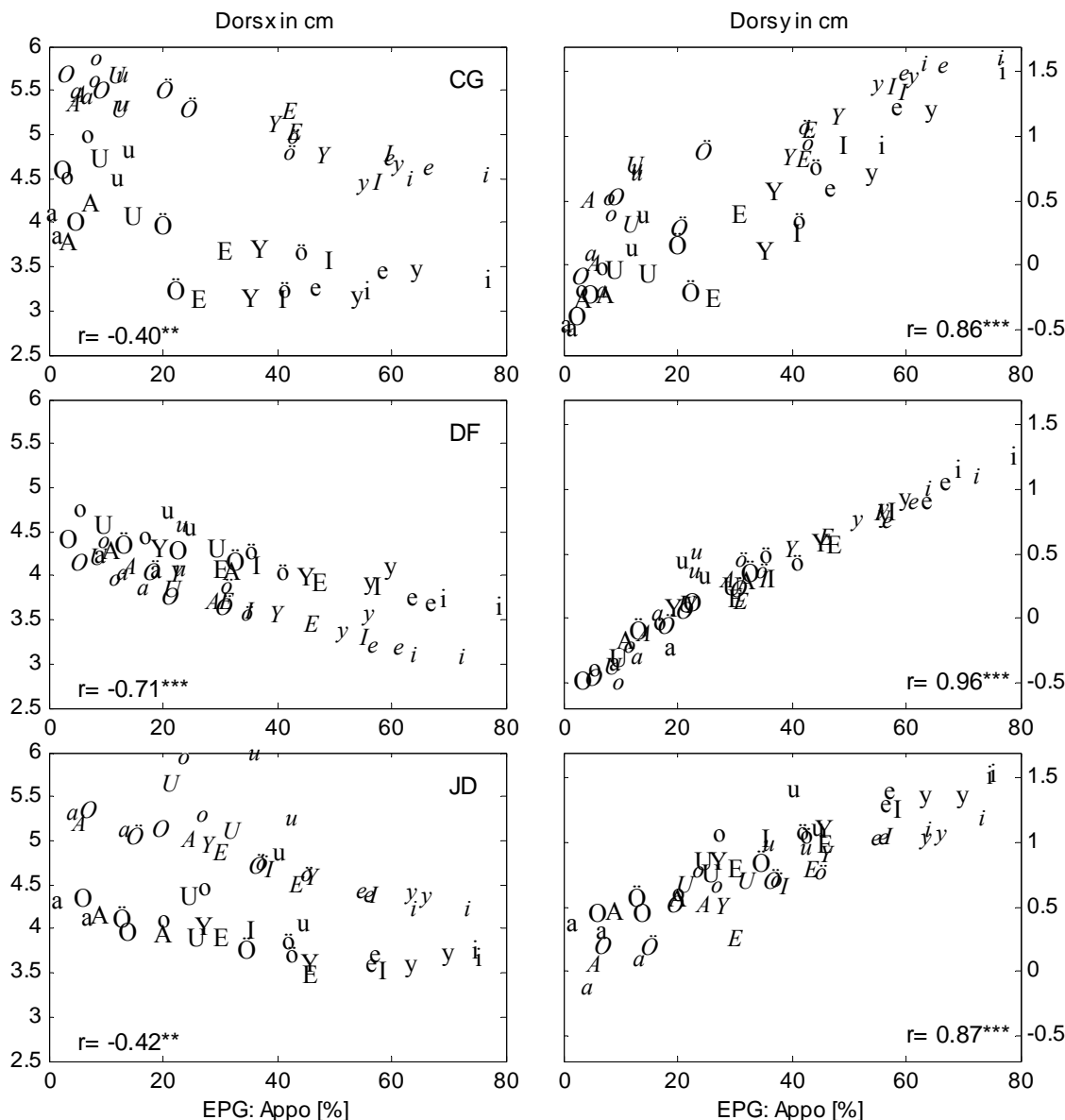
A number of conventional EPG measures was calculated, e.g. the centre of gravity, the percentage of contacts in the posterior region, and the centrality index (see e.g. Gibbon & Nicolaidis 1999). Since none of these measures account for either the non-uniform spatial distribution of electrodes on the artificial palate, or for individual differences in the size of the palate, the spacing between electrodes is not considered in the above mentioned measures. Hence a new approach was adopted by measuring the EPG 3D coordinates by a caliper on the individual artificial palates (for a detailed description see Fitzpatrick and Ní Chasaide 2002) and then computing the area around the contacts by triangulation. The new EPG index APPOPC (Area of Posterior Palatal Contact in Percent) was calculated as the percentage of areas around the activated contacts in the posterior region divided by the whole posterior area.

## 3. Results

### 3.1. *Relationship between tongue positions and palatal contact*

The two methods used here for recording lingual articulation can be seen as complementary for consonants. Indeed, the EPG electrodes detect whether there is a contact between tongue and palate and they describe the 3D distribution of these contacts, but they do not give any information about the tongue shape outside of the contact region. EMMA, on the other hand, provides accurate data about the position of the anterior part of the tongue in the mid-sagittal plane, but does not inform about the position of the sides of the tongue.

One of the aims of the current study is to investigate the influence of the amount of palatal contacts on the patterns of token-to-token variability in vowel production. Therefore as a first step, the relationship between the horizontal and vertical position of the tongue sensors and the EPG measure APPOPC was analysed by calculating correlation coefficients. This is important especially since both EPG and EMMA are limited spatially to the hard palate and to the anterior part of the tongue. Table I shows the correlation coefficients between horizontal and vertical sensor positions and the EPG area index APPOPC averaged over the 10-11 repetitions of each item split by condition. Figure 1 shows the corresponding scatterplots for the tongue dorsum position.



**Figure 1.** Scatterplots of the averaged horizontal (left column) and the vertical tongue dorsum positions (right column) and the EPG measure area of palatal contact in the posterior region in percent (APPOPC), shown by row separately for the three speakers. Upper case letters: lax vowels ( $\ddot{O}$ =/œ/), lower case: tense vowels ( $\ddot{o}$ =/ø/), italics: bite-block condition.

For all subjects significant negative correlations were found between APPOPC and the horizontal positions of the three tongue sensors, i.e. the further forward the vowel was produced the more EPG contacts were found. The relationship between the amount of EPG contact and vertical tongue sensor positions was also highly significant: higher tongue positions yielded an increase in the amount of EPG contact. For all speakers, the correlations were weaker for the tongue back sensor as compared to the two more anterior sensors. This could be due to the fact that the tongue back sensor was placed behind the posterior border of the artificial palate and therefore only indirectly contributed to the amount of measured palatal contact. As can be seen in Figure 1, when all conditions were considered together, speakers CG and JD showed correlations that were considerably lower for the horizontal dimension than for the vertical one. However, for these speakers the lower correlation coefficients with

horizontal positions were due to quite different tongue sensors locations for the recordings with and without bite-block as Table I shows: here the significance of correlation coefficients did not differ for horizontal and vertical positions when computed for the two bite-block conditions. Nevertheless, as mentioned above, the constriction location for high back vowels is very likely to be located behind the posterior border of the artificial EPG palate. Hence, since the real amount of contacts for these vowels was probably quite inaccurately measured with the EPG System the magnitude of the correlation along the horizontal axis should be considered with some caution.

TABLE I: Correlation coefficients between tongue positions of TBACK, TDORS and TBLADE and APPOPC, averaged over the 10-11 repetitions and split by condition: B=bite-block, N=normal. All correlations are significant, values in italics are significant at  $p<0.01$ , all others at  $p<0.001$

| Speaker | BB | N  | BACKX         | BACKY | DORSX         | DORSY | BLADEX        | BLADEY |
|---------|----|----|---------------|-------|---------------|-------|---------------|--------|
| CG      | B  | 28 | -0.853        | 0.645 | -0.924        | 0.923 | -0.968        | 0.965  |
|         | N  | 28 | <i>-0.514</i> | 0.861 | <i>-0.707</i> | 0.928 | <i>-0.798</i> | 0.827  |
| DF      | B  | 28 | -0.866        | 0.839 | -0.891        | 0.954 | -0.928        | 0.820  |
|         | N  | 28 | <i>-0.762</i> | 0.853 | <i>-0.849</i> | 0.972 | <i>-0.886</i> | 0.843  |
| JD      | B  | 28 | -0.689        | 0.677 | -0.740        | 0.906 | -0.791        | 0.887  |
|         | N  | 28 | <i>-0.598</i> | 0.817 | <i>-0.609</i> | 0.947 | <i>-0.780</i> | 0.870  |
| All     | B  | 84 | -0.423        | 0.602 | -0.459        | 0.833 | -0.723        | 0.704  |
|         | N  | 84 | <i>-0.475</i> | 0.560 | <i>-0.603</i> | 0.840 | <i>-0.767</i> | 0.763  |

In summary, our data have confirmed what was expected: the amount of palatal contact area increased for higher and more fronted tongue positions.

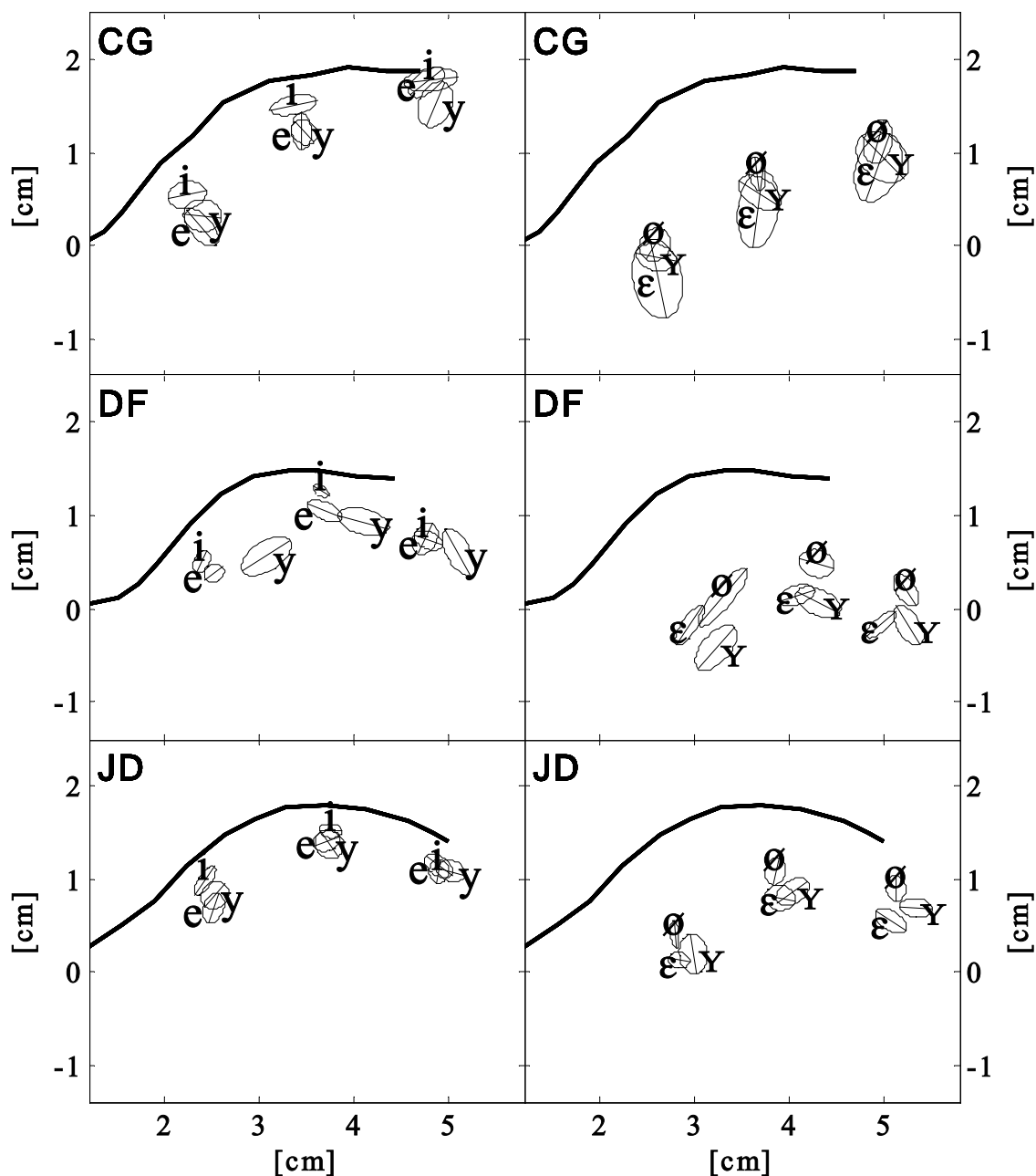
### 3.2. Ellipse areas and palatal contacts

Following the predictions of Perkell & Nelson (1984), we measured the relationship between amount of palatal contact and patterns of variability. More specifically, our hypothesis is that the higher the amount of contacts, the smaller the ellipse area. Examples of ellipses for selected vowels are given in Figure 2.

As can be seen in Table II, there was a close relationship between the ellipse area and the amount of palatal contact, i.e. the more the tongue was constrained by the palate the smaller the ellipse area. This relationship was strongest for the tongue dorsum sensor, which was located in the constriction region of the front vowels. It was clearly weaker for the TBACK sensor, here again probably because it was placed behind the posterior border of the EPG palate. Thus, this preliminary result tends to support our first hypothesis that articulatory variability should decrease when the amount of palatal contacts increases.

TABLE II: Correlation coefficients between EPG-measure APPOPC and ellipse areas (italics:  $p<0.05$ , bold:  $p<0.01$ ).

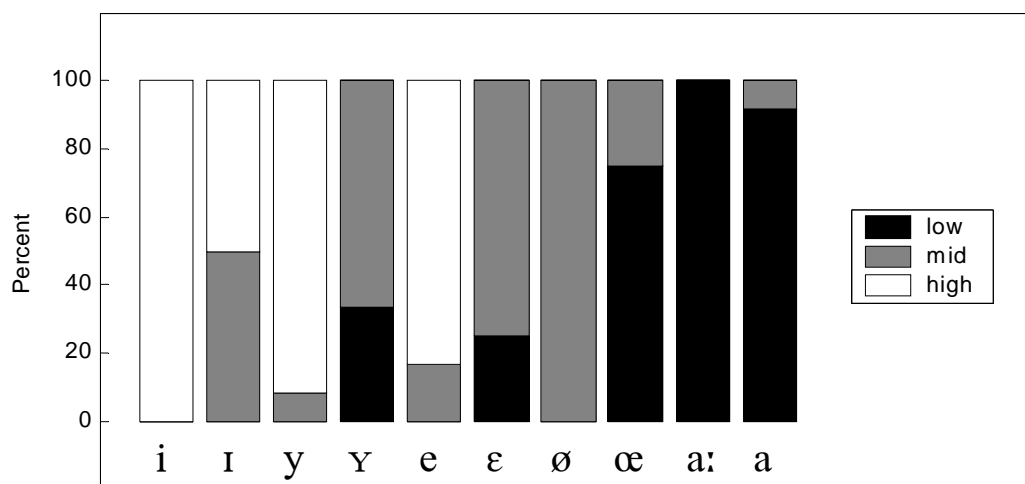
| Speaker | N   | Area Back     | Area Dors     | Area Blade    |
|---------|-----|---------------|---------------|---------------|
| CG      | 56  | -0.115        | <b>-0.386</b> | <i>-0.275</i> |
| DF      | 56  | -0.192        | <b>-0.345</b> | <i>-0.252</i> |
| JD      | 56  | <i>-0.299</i> | <b>-0.419</b> | <i>-0.317</i> |
| All     | 168 | <b>-0.209</b> | <b>-0.379</b> | <b>-0.296</b> |



**Figure 2.**  $2\sigma$  ellipses for selected vowels in bilabial context without bite-block for the three speakers. Panels on the left show ellipses for high front vowels and panels on the right for mid front vowels of the three sensors tongue back, tongue dorsum and tongue blade.

However, for each speaker, the number of articulatory tokens measured for each vowel and for each of the four conditions (2 consonantal contexts, and 2 bite-block conditions) varied between 8 and 11. For such a small amount of data, the reliability of the ellipse areas is questionable. And, as a matter of fact, we have stated experimentally that under such conditions the orientation of the ellipse was very sensitive to the presence of one or two possible outliers. Hence, in order to provide a reliable analysis of the relations between amount of palatal contacts and data dispersion, we built up larger data sets that pooled separately for each speaker and for each of the 4 conditions, data from different vowel categories. The obvious method to this is to group vowels together according to their phonological features such as vowel height and frontness. We rejected this method for two reasons: first of all, speakers varied in their relative target position of specific vowels, e.g.

speaker DF produced the vowel /y/ lower and more retracted than the other speakers (see Figure 2). Secondly, the articulatory positions of vowels varied with the consonantal context and the bite-block condition. Therefore pooling vowels together based on phonological features would be rather subjective and a priori. An alternative and more objective method is to transform a continuous variable into a categorical. Since we are looking for a possible influence of the amount of palatal contacts on the orientation of the dispersion, the new data subsets were built for each speaker and for each of the four conditions on the basis of the EPG parameter APPOPC. For further analysis, the function “categorize variable” of the statistics software SPSS™ was used. According to percentiles of the continuous variable APPOPC, this function computes a predefined number of levels that define the different data subsets. For example, if the new categorical variable consists of three levels, the SPSS procedure assigns 1 to all data of the original variable which are smaller than the 33<sup>rd</sup> percentile. These levels define a new, discrete, variable, called NAPPOPC that specifies each of the new data subsets. It should be noted that back vowels /u, ʊ, o, ɔ/ were excluded from this automatic categorization and that we arbitrarily grouped them in a subset called “back”. We did this because the constriction location of these vowels is just at or even behind the posterior border of the EPG palate, so that the amount of contact at the soft palate cannot be reliably measured. Therefore, the relationship between EPG contact patterns and lingual articulation differs considerably from that of the front vowels, and no relevant conclusion about the effect of palatal contact on lingual variability can be made for these 4 vowels. This is why they were grouped together in a separate data subset. For the other vowels, we decided to construct 3 categories.

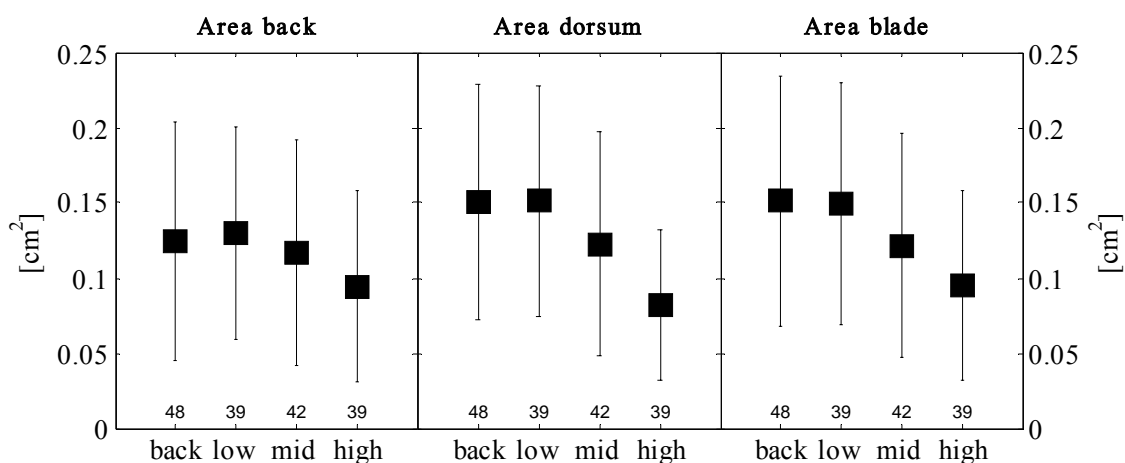


**Figure 3.** Distribution of the different vowel categories among the three new data subsets derived from the three-level quantization of the continuous variable APPOPC (all back vowels are excluded). Y-axis : Percentage of vowels distributed in the low (black), mid (gray), high (white) categories

Figure 3 shows how the different repetitions of each vowel were distributed among the 3 categories pooled for all speakers and conditions. This chart was generated by counting the occurrences of categories 1, 2 and 3 per vowel type and calculating the percentage. The maximum number for each vowel type was 12 (3 speakers, 2 consonants and 2 conditions) and for each vowel the APPOPC value was averaged over the 8-11 repetitions. It can be seen that the high front vowels /i, y, e/ were generally grouped together, henceforth called "high" (white bars), and that the majority of the low vowels /a:, a/ were in the same category, henceforth called "low" (black bars). All 12 instances of the tense front vowel /ø/ received a

medium value, henceforth called "mid" (gray bars). The other vowels were distributed less consistently depending on speaker, consonantal context and bite-block condition.

Then, for each of the new data subsets, back, low, mid and high, the ellipse areas were computed for the three tongue sensors and they were averaged across all speakers. Figure 4 shows the results. They were consistent with the findings made for each vowel category separately (see Table II). Indeed, vowels with a high amount of palatal contact generally exhibited less token-to-token variability than back or low vowels. Ellipse areas of vowels with an intermediate level of palatal contact were between high and low or back vowels. Low and back vowels did not differ in their amount of variability, but it should be recalled that measurements for back vowels have to be interpreted with caution.



**Figure 4.** Means and standard deviations of ellipse areas in  $\text{cm}^2$  for the tongue back sensor (left), the tongue dorsum sensor (mid) and the tongue blade sensor (right) split by the four categories back, low, mid and high according to the amount of palatal contact. Data are averaged over all speakers. The numbers below indicate the number of ellipses per category.

To test whether these differences are significant, ANOVAs were computed with ellipse areas of the three sensors as dependent variable and the discrete variable NAPPOPC as the independent variable with the four levels “back”, “low”, “mid” and “high”. The upper part of Table III shows significant differences pooled over all speakers while the lower parts give significant differences for individual speakers.

Across all speakers, the areas of the tongue dorsum sensor measured for high vowels were systematically significantly different from the areas measured for the other subsets. However, this didn't apply to mid vowels, which did not significantly differ from back or low vowels. Therefore it could be assumed that the effect of palatal contact on the amount of token-to-token variability was restricted to high front vowels. For single speaker analyses, significant differences were found less often, they were generally weaker, and they did not always confirm the significant differences found for the whole set of speakers. Indeed, the general tendency of smaller ellipses for high vowels was only valid for CG. For speaker DF no significant differences were found. However, this can be explained probably by the fact that speaker DF in general exhibited very little token-to-token variability for all vowel categories. For speaker JD, the only significant difference was between high and back vowels, and as already mentioned results for the "back" category should be considered with caution. Hence, individual data only weakly support the hypothesis of a major reduction of variability associated with a large amount of contact. Our results suggest that such a relationship could be speaker dependent. When there is a tendency for a speaker to display a large token-to-

*Token-to-token Variability*

token variability, this variability is reduced when there is a large amount of contact between tongue and palate. Otherwise, the role of the palate is barely noticeable.

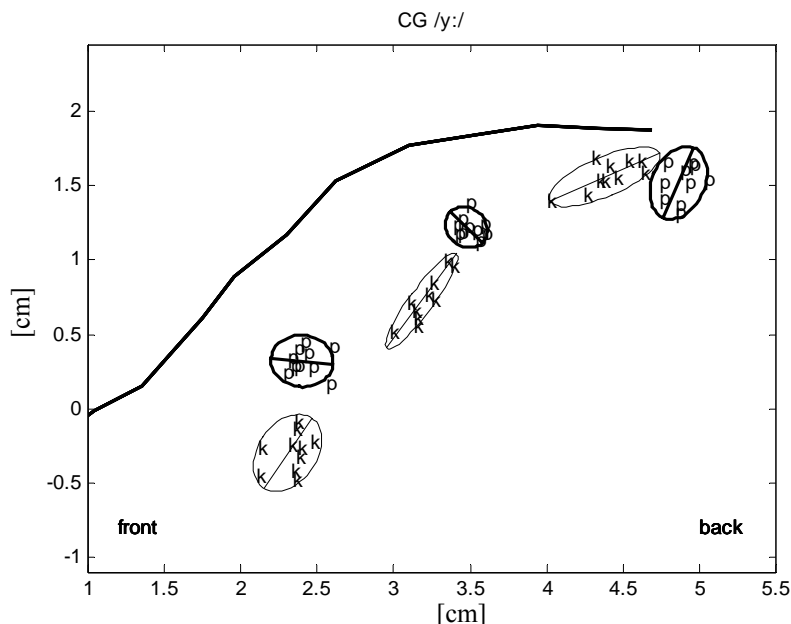
TABLE III. Results of ANOVAs with the dependent variable ellipse area for the three sensors and the independent variable NAPPOPC for all speakers pooled (ALL) and for the three speakers CG, DF, JD separately. Results of Post hoc Scheffé tests for the 4 levels of NAPPOPC are also given, e.g. an asterisk in row *back > high* means that ellipse areas for back vowels were significantly larger than for high vowels.

| Speaker |             |        | Back   |       | Dorsum |       | Blade  |       |  |
|---------|-------------|--------|--------|-------|--------|-------|--------|-------|--|
| ALL     | Effects     | df     | F      | p     | F      | p     | F      | p     |  |
| GLM     | NAPPOPC     | 3, 167 | 3.063  | 0.030 | 12.040 | 0.000 | 6.721  | 0.000 |  |
|         | Speaker     | 2, 167 | 52.951 | 0.000 | 34.928 | 0.000 | 29.348 | 0.000 |  |
|         | Interaction | 6, 167 | 0.798  | 0.573 | 0.700  | 0.650 | 2.093  | 0.057 |  |
|         | Post hoc    | back > | low    |       |        |       |        |       |  |
|         |             |        | mid    |       |        |       |        |       |  |
|         |             |        | high   |       |        | ***   |        | **    |  |
|         |             | low >  | mid    |       |        |       |        |       |  |
| high    |             |        |        | ***   |        | **    |        |       |  |
| mid >   | high        |        |        | *     |        |       |        |       |  |
| <hr/>   |             |        |        |       |        |       |        |       |  |
| CG      |             |        | Back   |       | Dorsum |       | Blade  |       |  |
| GLM     | Effects     | df     | F      | p     | F      | p     | F      | p     |  |
| GLM     | NAPPOPC     | 3, 55  | 1.003  | 0.399 | 4.674  | 0.006 | 1.946  | 0.134 |  |
|         | Post hoc    | back > | low    |       |        |       |        |       |  |
|         |             |        | mid    |       |        |       |        |       |  |
|         |             |        | high   |       |        | *     |        |       |  |
|         |             | low >  | mid    |       |        |       |        |       |  |
| high    |             |        |        | *     |        |       |        |       |  |
| mid >   | high        |        |        |       |        |       |        |       |  |
| <hr/>   |             |        |        |       |        |       |        |       |  |
| DF      |             |        | Back   |       | Dorsum |       | Blade  |       |  |
| GLM     | Effects     | df     | F      | p     | F      | p     | F      | p     |  |
| GLM     | NAPPOPC     | 3, 55  | 2.308  | 0.087 | 3.163  | 0.032 | 1.700  | 0.178 |  |
|         | Post hoc    | back > | low    |       |        |       |        |       |  |
|         |             |        | mid    |       |        |       |        |       |  |
|         |             |        | high   |       |        |       |        |       |  |
|         |             | low >  | mid    |       |        |       |        |       |  |
| high    |             |        |        |       |        |       |        |       |  |
| mid >   | high        |        |        |       |        |       |        |       |  |
| <hr/>   |             |        |        |       |        |       |        |       |  |
| JD      |             |        | Back   |       | Dorsum |       | Blade  |       |  |
| GLM     | Effects     | df     | F      | p     | F      | p     | F      | p     |  |
| GLM     | NAPPOPC     | 3, 55  | 2.397  | 0.079 | 7.267  | 0.000 | 12.478 | 0.000 |  |
|         | Post hoc    | back > | low    |       |        |       | **     |       |  |
|         |             |        | mid    |       |        | *     | ***    |       |  |
|         |             |        | high   |       |        | **    | ***    |       |  |
|         |             | low >  | mid    |       |        |       |        |       |  |
| high    |             |        |        |       |        |       |        |       |  |
| mid >   | high        |        |        |       |        |       |        |       |  |

One of the major aims of this study is to investigate not only the amount of token-to-token variability but also its orientation. Our hypothesis is that "for high vowels, which have a high amount of palatal contacts, the major axis of the dispersion ellipses should be oriented along the vocal-tract midline", i.e. along the outline of the palate.

Figure 5 shows the ellipses of /by:pə/ and /gy:kə/ for tongue blade, tongue dorsum and tongue back sensors for speaker CG without bite-block. As discussed earlier, both the shape and the orientation of the ellipses were highly sensitive to single outliers. For instance, without the lower outlier of the tongue blade measurements for the bilabial context the front ellipse (bold) would be oriented along the palate and it would be flatter instead of being nearly circular and inclined almost parallel to the x-axis. We therefore once again decided to consider larger data subsets by grouping vowels categories together on the basis of the

amount of palatal contacts by using the three-level categorization of APPOPC (which specifies the discrete variable NAPPOPC, see Figure 3).



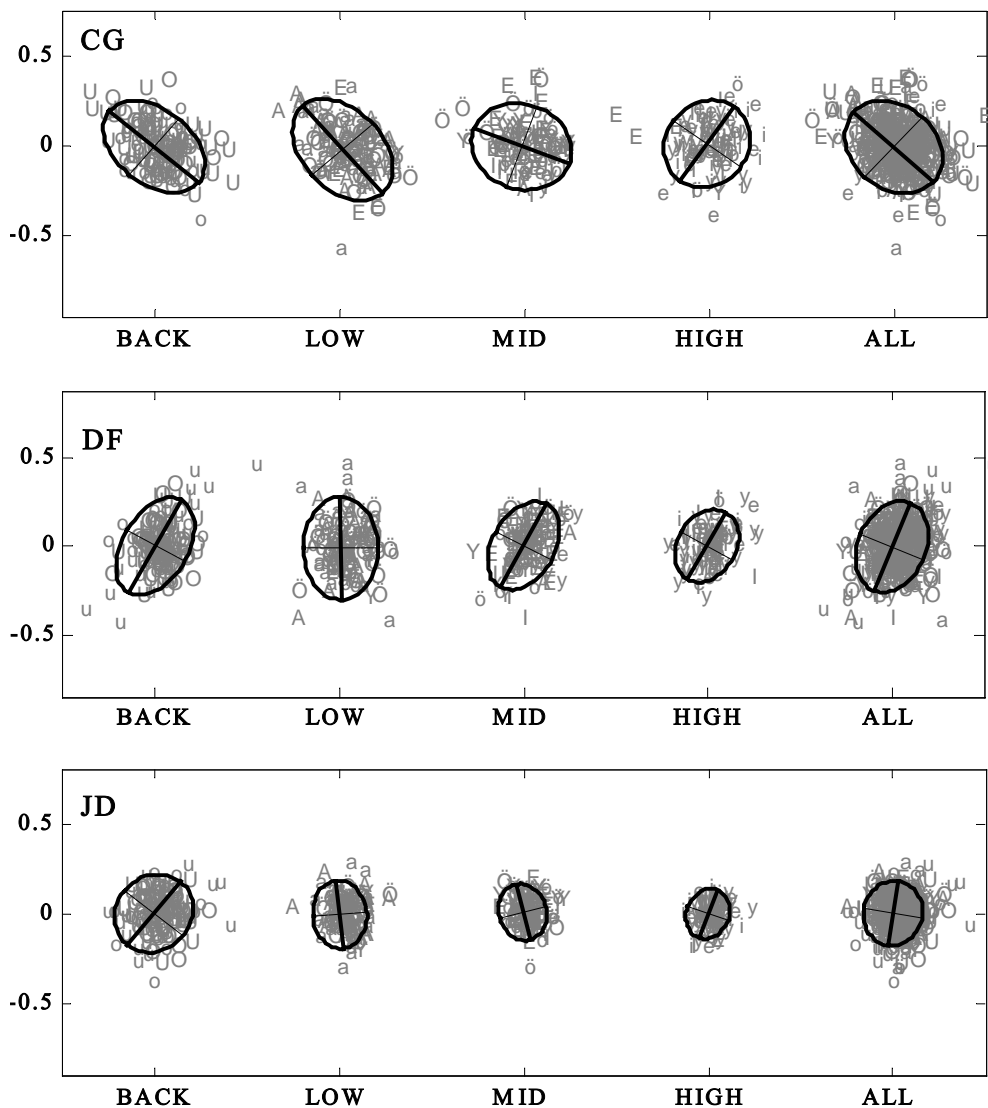
**Figure 5.**  $2\sigma$  ellipses of 10 repetitions of /by:pə/ (bold lines, items marked with p) and /gy:kə/ (thin lines, items marked with k) of sensor positions of tongue blade, tongue dorsum and tongue back for speaker CG without bite-block, measure approximately at the mid-vowel.

TABLE IV: Correlation coefficients between the x and y positions within the subsets Back, Low, Mid, High and for all measures. Gray cells correspond to significant correlations ( $p < 0.05$ ) and bold font is used if  $p < 0.001$

| Speaker |        | Back          | Low           | Mid           | High          | All           |
|---------|--------|---------------|---------------|---------------|---------------|---------------|
| CG      | N      | 156           | 133           | 127           | 130           | 546           |
|         | Tback  | <b>-0.474</b> | -0.218        | -0.228        | 0.016         | <b>-0.252</b> |
|         | Tdors  | <b>-0.409</b> | <b>-0.324</b> | <b>-0.282</b> | <b>0.230</b>  | <b>-0.232</b> |
|         | Tblade | <b>-0.394</b> | <b>-0.373</b> | -0.104        | 0.114         | <b>-0.226</b> |
| DF      | N      | 163           | 131           | 140           | 130           | 564           |
|         | Tback  | <b>-0.552</b> | -0.182        | <b>-0.384</b> | <b>-0.444</b> | <b>-0.392</b> |
|         | Tdors  | <b>-0.218</b> | -0.040        | -0.091        | <b>-0.422</b> | <b>-0.172</b> |
|         | Tblade | <b>0.403</b>  | -0.003        | <b>0.345</b>  | <b>0.237</b>  | <b>0.250</b>  |
| JD      | N      | 160           | 131           | 136           | 130           | 557           |
|         | Tback  | <b>-0.260</b> | <b>-0.352</b> | <b>-0.300</b> | 0.083         | <b>-0.239</b> |
|         | Tdors  | 0.018         | -0.222        | -0.154        | 0.052         | -0.082        |
|         | Tblade | 0.111         | -0.050        | -0.098        | 0.126         | 0.038         |

Before calculating ellipse orientations and sizes for the larger data subsets, the data were centered. To do so, for each sensor, the X and Y mean values were calculated for each individual vowel category, each consonantal context and each bite-block condition separately, and they were subtracted from the original corresponding data. The centered data were then grouped according to the discrete variable NAPPOPC. Table IV gives the correlation coefficients between the x and y positions within the subsets Back, Low, Mid, High and pooled (All). A significant negative correlation means that the higher the tongue the more

fronted it is. It also means that for the sensors located in the palato-alveolar region (i.e. for the tongue blade sensor and in some cases for the tongue dorsum sensors, see Figure 2), which is where the palate outline goes down when it goes forward, the variation is mainly perpendicular to the palate outline. A positive correlation means that the higher the tongue the more retracted it is. In such a case, it can be concluded that the main orientation of the variation of the tongue sensors located in the palato-alveolar region is parallel to the palate outline. Significant positive and negative correlations imply flat ellipses. When there is no significant correlation between x- and y-positions then the ellipses will be more or less circular.



**Figure 6.** Scatter plots with superimposed  $2\sigma$  ellipses with the major axis marked for the tongue dorsum sensor grouped according to the amount of palatal contact (BACK, LOW, MID and HIGH) and overall vowels (ALL) in cm. For graphical reasons tick marks are not shown for the x-axis but the scaling of x and y-axes are set equal. All data are centralized to group means.

As can be seen in Table IV and in Figure 6, there was no consistent pattern across speakers. Speaker CG shows significant negative correlations for all sensors for the vowel subsets Back, Low and Mid (except for Tblade). For High vowels, correlations were either non significant (Tback, Tblade) or positive. This is in general agreement with our

hypotheses: the main orientation of the articulatory variability was different when a large amount of contact between the tongue and the palate existed, and in this case the variability was mainly limited in the direction orthogonal to the palate. Speaker DF presented a less clear picture: the main orientation of the variability varied with the position of the sensor on the tongue, and it did not seem to depend on the vowel subset: for this speaker the large majority of the correlations were significant, and systematic negative correlations were observed for the Tback and Tdors sensors, while correlations were positive for the Tblade sensor, when they were significant. A third pattern was observed for speaker JD: the majority of the correlation were not significant, and when they were significant, the correlations were always negative, which was not in agreement with our hypotheses.

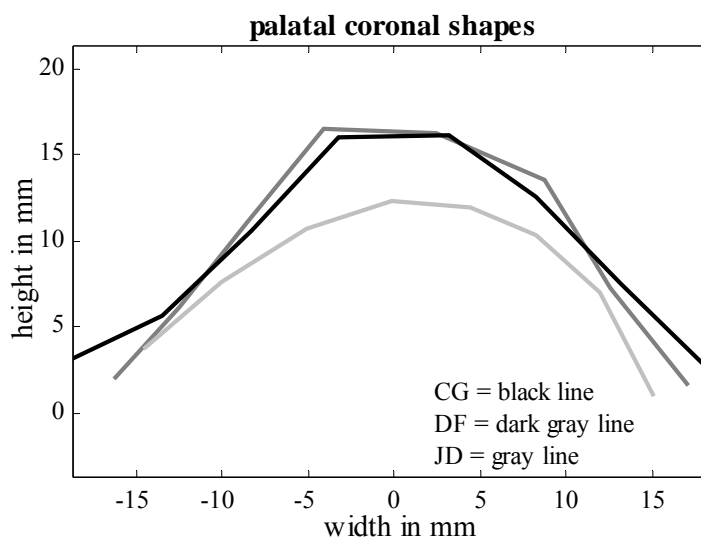
#### 4. Discussion and Conclusions

For high vowels, our results do not confirm Perkell & Nelson's (1985) or Beckman *et al.*'s (1995) observations for American English speakers: while these two studies found for all subjects flat dispersion ellipses oriented along the palate outline for high front vowels, only one of our speakers, speaker CG, showed the same trend. Based on our findings, Perkell & Nelson's (1985) hypothesis that the variability for high vowels is constrained by tongue-palate contacts to be orthogonally inclined to the palate, does not generally apply for German subjects. However, this conclusion should be tempered because the data in Figure 6 also show that for speaker JD the high vowels' articulatory variability is less important than for DF and clearly less important than for CG. This suggests that for some reason the production of high vowels could be much more constrained for JD than for DF and CG. Two factors could explain this tighter constraint.

First of all, the vowel inventory in German is much more crowded than in English. For example, if we only consider the high front vowels, German has three phonemes that are in very close proximity to each other, namely /i, y, e/ (German /e/ is a phonetically closer vowel than the cardinal vowel 2). Consequently the articulation of these sounds has to be very accurate and different strategies could be used by the different speakers. As can be seen in Figure 2, a high amount of overlap was found for speaker CG between /e/ and /y/. The speaker achieved a very clear perceptual differentiation between these vowels using lip-rounding with a mean difference in horizontal lip position between /e/ and /y/ of 1.27 cm in the velar context and 0.83 cm in the bilabial context. In contrast, since speaker DF exhibited almost no lip protrusion differences (bilabial: 0.11 cm, velar: 0.24 cm), he had to adopt a retracted tongue position for /y/ in all four conditions to maintain the perceptual distinctiveness. As far as /i/ and /e/ are concerned, their distinctiveness can obviously not be enhanced by these types of compensatory lip-tongue strategies (since they are both unrounded vowels), and because the German vowel inventory is more crowded than in English, the tongue position variability for these sounds is likely to be much smaller in German than in their English counterparts. If the variability is constrained to be less, then the size of the major axis will be correspondingly less important and the ellipse will be closer to a circle, which makes the detection of its main orientation more difficult and then more variable.

Hence, differences in the density of the vowel systems seem to explain, at least in part, the differences between the extent of variability observed for our German subjects and that measured by Perkell & Cohen (1989) for native speakers of English. However, even among the German subjects, the measured patterns are quite different and the following question remains unresolved: why are these gesture accuracy requirements stronger for JD than for DF and for CG? This speaker neither uses lip protrusion (the difference between /y/ and /e/ is about 3 mm and comparable to speaker DF's) nor tongue retraction for /y/ but positions his tongue with an extremely high precision. An explanation can be found by observing the

coronal shape of the palate of each speaker in the region of the vocal tract where the cross-sectional area reaches its minimum (constriction location of the vowel). Speaker JD (gray line) has a palate that is much flatter and wider in its upper part than the palate of speakers CG and DF. Consequently, for high vowels a given vertical displacement of the tongue is likely to induce for this speaker an increase of the cross-sectional area at the constriction, which is clearly larger than for speakers DF and CG, and this will in turn induce larger acoustic changes (see Majid et al., 1987 or Gay et al., 1991). This hypothesis is in agreement with Perkell *et al.* (1997) who provided evidence of a relation between the amount of token-to-token articulatory variability for /u/ and the coronal shape of the palate. These authors show that one of the subjects, who has a broad and flat palatal shape, exhibits a small variability with dispersion ellipses nearly circular. The second subject with a much narrower palatal shape shows a rather large variability with dispersion ellipses having their main orientation parallel to the palatal outline. In our data, the palates of subjects CG and DF have similarities with the palate of the second subject of Perkell et al.'s (1997) data, while speaker JD presents similarities with their first speaker. Hence, the nearly circular shape of the dispersion ellipses and the smallest articulatory variability observed for speaker JD could be due, at least for high vowels, to his palatal shape.



**Figure 7.** Coronal shapes of the EPG palates of the speakers (at the 2<sup>nd</sup> last most posterior row = Constriction region).

Two conclusions can be drawn from these results: First, not only contextual variability as was found by Manuel (1990), but also the amount and pattern of token-to-token variability is likely to be constrained by the density of the vowel inventory. More precisely, the high front vowels in German are produced with tongue positions that are very close to each other. In contrast, American English which was studied in Perkell and Nelson and in Beckman et al. has only a single vowel in this region. We can therefore conclude that even if there might be an influence of the amount of palatal contact by limiting the variability in the direction of the palate outline, the variability in the direction of constriction location is further constrained by language-specific characteristics such as the vowel inventory.

A second conclusion is that speakers differ with respect to the strategies they use to maintain the perceptual distinction between categories (e.g. lip rounding for speaker CG vs. tongue retraction for speaker DF). The reasons for developing different strategies could lie in individual morphology as was found in our data for palate shape (Speakers CG, DF vs. JD).

However, we are still without an explanation for why the orientation in DF's high vowel ellipse does not conform to the predictions about the role of the tongue-palate interaction.

We intend to assess the discrepancy between our results and the original hypotheses (see the end of section 1) by making use of 2D and 3D biomechanical models of the tongue (Perrier et al., 2003; Gerard et al., 2003). The motor commands will be corrupted by different kinds of noise and the corresponding articulatory variability will be analyzed for different conditions. First, different palatal shapes will be considered in the midsagittal plane (steep versus flat in the alveolar region, simulations made) and in the coronal plane (wide versus narrow in its upper part). This will allow us to assess the role of the palatal geometry in the articulatory variability patterns. Second, different muscle activities will be considered for each of the vowels; indeed it is well-known that due to synergies and antagonisms between muscles, very similar tongue shapes can be produced with different muscle recruitments, and, as a result, various levels of force. This should enable us to test the possible consequences for articulatory variability of the different strategies elaborated by speakers in maintaining the perceptual distinctiveness according to their vocal tract geometry or the density of their vowel system. Finally, different noise amplitudes will be considered in order to observe how variability patterns vary, when the amount of noise in the muscle-activity level changes.

This work was supported by a grant from the German Research Council (DFG) GWZ 4/8-1, P.1. We also like to thank Jörg Dreyer from the ZAS for technical support during the EMMA and EPG recordings and Jonathan Harrington for very valuable comments.

## References

- Beckman M.E., Jung T.-P., Lee S.-L., de Jong K., Krishnamurthy A.K., Ahalt S.C., Cohen K.B., & Collins M.J. (1995) Variability in the production of quantal vowels revisited. *Journal of the Acoustical Society of America*, **97**, pp. 471-490.
- Blumstein S.E. & Stevens K.N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, **66** (4), pp. 1001-1017.
- Blumstein S.E. & Stevens K.N. (1980). Perceptual invariance and onset spectra for stop consonants in various vowel environments. *Journal of the Acoustical Society of America*, **67**, pp. 648-662.
- Browman C.P. & Goldstein L.M. (1985). Dynamic modeling of phonetic structure. In *Phonetic linguistics* (V. Fromkin, editor) pp. 35-53 New York: Academic.
- Browman C.P. & Goldstein L.M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, pp. 219-252.
- Browman C.P. & Goldstein L.M. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, pp. 299-320.
- Fitzpatrick, L. & A. Ní Chasaide (2002). Estimating lingual constriction location in high vowels: a comparison of EMA- and EPG-based measures. *Journal of Phonetics* **30**, pp. 397-415.
- Folkins J.W & Brown C.K. (1987). Upper lip, lower lip and jaw interactions in speech: comments on evidence from repetition-to-repetition variability, *Journal of the Acoustical Society of America*, **82**, pp. 1919-1924.
- Fujimura O. (1986). Relative invariance of articulatory movements : An Iceberg Model. In *Invariance & Variability in speech processes* (J.S. Perkell & D.H. Klatt, editors) pp. 226-234. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Gay T., Boë L.J., Perrier P., Feng G. & Swayne E. (1991). The acoustic sensitivity of vocal tract constrictions : a preliminary report. *Journal of Phonetics*, **19**, pp. 445-452.
- Gérard J.M., Wilhelms-Tricarico R., Perrier P. & Payan Y. (2003) A 3D dynamical biomechanical tongue model to study speech motor control. *Recent Research Developments in Biomechanics*, **1**, pp. 49-64, Transworld Research Network.
- Gibbon, F. & K. Nicolaidis (1999) Palatography. In *Coarticulation* (W. J. Hardcastle & N. Hewlett, editors), pp. 229-245. Cambridge: University Press.

## Token-to-token Variability

- Guenther, F. H. (1995). Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychological Review*, **102**, pp. 594–62.
- Guenther, F. H., Hampson, M. & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, **105**, pp. 611–633.
- Guenther, F. H., Espy-Wilson, C.Y., Boyce, S.E., Matthies, M.L., Zandipour, M. & Perkell, J.S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, **105**, 2854–2865.
- Harris, C.M. (1998). On the optimal control of behaviour: a stochastic perspective. *Journal of Neuroscience Methods*, **83**, pp. 73-88.
- Harris, C.M. & Wolpert, D.M. (1998). Signal dependent noise determines motor planning. *Nature*, **394**, pp. 780-784.
- Keating P.A. (1990) The window model of coarticulation: articulatory evidence, In *Papers in Laboratory Phonology I* (J. Kingston & M. Beckman, editors) pp. 451-470, Cambridge: University Press.
- Kuhl P. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, **50**, pp. 93-107.
- Kuhl P. (1992). Infants perception and representation of speech: development of a new theory. *Proc. of the 2<sup>nd</sup> International Conference of Spoken Language Processing* pp. 449-456. University of Alberta, Canada.
- Lindblom B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In *Speech production and speech modelling* (W.J. Hardcastle & A. Marchal, editors), pp. 403-439. Dordrecht: Kluwer Academic Publishers.
- Lubker J. & Gay T. (1982), Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America*, **71**, pp. 437-448
- Majid R., Abry C., Boë L.J. & Perrier P. (1987). Contribution à la classification articulatoire-acoustique des voyelles : étude des macro-sensibilités à l'aide d'un modèle articulatoire. *Proceedings of the 11th International Congress of Phonetic Sciences* **2**, pp. 348-351. Tallin, Estonia
- Mooshammer C., Perrier P., Fuchs S., Geng C. & Payan P. (2001). The control of token-to-token variability: an experimental and modeling study. *Actes de la 4th International Speech Motor Conference*, pp. 78-81, Nijmegen.
- Perkell J.S. & Nelson W.L. (1985). Variability in production of the vowels /i/ and /a/, *Journal of the Acoustical Society of America*, **77**, pp. 1889-1895.
- Perkell J.S. & Cohen M.H. (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics*, **17**, pp. 123-133.
- Perkell J.S. (1990) Testing theories of speech production: implications of some detailed analyses of variable articulatory data. In *Speech Production and Speech Modelling* (W.J. Hardcastle and A. Marchal, editors) pp. 263-288. Dordrecht: Kluwer Academic Publishers.
- Perkell, J., Matthies M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication*, **22**, pp. 227-250.
- Perkell J.S., Guenther F.H., Lane H., Matthies L.M., Perrier P., Vick J., Wilhelms-Tricarico R., & Zandipour M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, **28** (3), pp. 233-272.
- Perrier P. (2003). About speech motor control complexity. *Proc. of the 6th International Seminar on Speech Production*, pp. 225-230, Sydney, December 7-10.
- Perrier P., Payan Y., Zandipour M. & Perkell J. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, **114** (3), pp. 1582-1599.
- Shiller D.M., Laboissière R. & Ostry D.J. (2002). Relationship between jaw stiffness and kinematic variability in speech. *J. Neurophysiology*, **88**, pp. 2329-2340
- Stevens K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human Communication: A unified view* (David Jr. E.E. & Denes P.B., editors), pp. 51-66. New York: Mc Graw Hill.
- Stevens K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, **17**, pp. 3-45.
- Stevens K.N. & Blumstein S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* **64**, pp. 1358-1368.
- Wood S. (1979). A radiographic examination of constriction location for vowels. *Journal of Phonetic*, **7**, pp. 25-43.



