

Some aspects of position and style in German
function words

Thomas Wesener

12 March 2001

Contents

1	Introduction	1
2	Method	2
3	The impact of position on personal pronouns	3
3.1	/r/-vocalization in <i>wir</i>	4
3.2	Vowel shortening in <i>ich</i>	16
3.3	Centralization in <i>Sie</i>	22
3.4	<i>es</i> and <i>'s</i>	23
3.5	Outlook on new scenario: <i>er</i>	28
4	Other phenomena in read speech	29
4.1	/x/	29
4.2	/h/ in read speech	36
4.3	Vowel nasalization	37
5	Discussion	37
A	Composition of the databases	41

1 Introduction

This paper builds on work on the reduction of function words in German spontaneous speech presented in Wesener (1999). Whereas the latter focused on one-word items and their reduction, we now aim to shed further light on reduction in sequences of items. The main interest will lie in the reduction of personal pronouns depending on the position relative to a preceding or following auxiliary verb. Hypotheses in Kohler (1979) and Kohler (1990) are the starting point. They e.g. state that the vowel of *ich* is likely to be dropped in *ich weiß*, but not in *weiß ich*. This study aims to broaden the basis for the investigation of these hypotheses by using computerized corpora of both spontaneous and read speech (IPDS 1995, IPDS 1996, IPDS 1997; IPDS 1994).

One might regard reduced forms of function words as clitics because are unlikely to occur in one-word utterances. But by contrast to typical clitics such as Latin *-que* ‘and’ in *arma virumque* (lit. ‘arms man-and’: ‘arms and the man’, cf. Matthews 1997), the function word investigated here can occur on their own. In addition, there frequently is no one-to-one relation between the reduced form and the item to which it is attached. The reduced function

words frequently occur in longer sequences containing other function words so that several ‘clitics’ refer to one word. Inversely, there may be two alternative words of reference for a reduced form, i.e. the preceding and the following. In this case it seems somewhat arbitrary to chose one word as the point of reference for the ‘clitic’. We therefore prefer the terminology ‘preverbal’ and ‘postverbal’ to ‘proclitic’ and ‘enclitic’, respectively. This does not exclude the possibility that certain pre- or postverbal reductions may be lexicalized (see discussion).

A survey in Wesener (1999) shows that first person pronouns are particularly frequent. They are thus ideal objects for investigating repetitions of the same pronoun-verb/verb-pronoun sequences. Third person pronouns, on the other hand, are rare because of the limitations of the scenario used to elicit the material. It is therefore difficult to investigate reduction of these pronouns induced by positional factors.

2 Method

The data derive from the sessions investigated in Wesener (1999) plus five isolated test dialogues published on the *Kiel Corpus* CDROMs. The *Kieldat* utility (Pätzold 1997) was used to create databases of the material. One database contains all prosodically and segmentally labelled material, another the data with segmental labels only, and a third the complete spontaneous material. These databases represent the status of January 2000 and are identical to those used in Rodgers (2000) (cf. appendix A).

Awk scripts were used to retrieve the sequences of interest, to count incidences of phenomena captured by the labelling, and to obtain durational information. The latter was retrieved from the prosodically labelled database, since sentence-accent with its strong durational cues is consistently marked in this material only. These automatically generated data were complemented by an impressionistic investigation of the sequences to capture phenomena which had not been labelled.

The ESPS tool **formant** was used to measure formant frequencies. *Formant* high-pass filters the input signal, down-sampled from 16 kHz to 10 kHz, to remove low frequency rumble (cut-off at approximately 80 Hz). The resulting file is then used for the formant frequency estimates (default values: window duration 49 ms, window type *Cos*⁴, preemphasis constant 0.7, LPC order 12, LPC type autocorrelation). It turned out that **formant** often has difficulties with identifying F1, F2 or F3 and chooses the next higher formant instead, tracking the wrong formant throughout the vowel. This problem could not be solved by changing window type, LPC order or other

factors.

String information from *Kieldat* databases was used to access signal files. The tools *klara* and *ksort* (Willems 1987, Scheffers and Simpson 1995) were then applied to estimate formant frequencies (window size 30 ms, Hamming window, preemphasis factor 0.95, LPC order 16). The programs seem to work more reliably for F1 and F2 by combining root solving of the LPC polynomial with the computation of Pisarenko frequencies (cf. Scheffers and Simpson 1995, Delsarte and Genin 1986). When problems occur with measuring higher formants, this is often indicated by nonsense values which are easy to detect. All unstressed renditions in the prosodically labelled database without hesitational lengthening were investigated. A script then calculated mean formant values from the retrieved data sets.

The frequency values were converted to perceptual units since Hz-values do not indicate whether changes are relevant in everyday communication and in the judgment of quality changes by phoneticians (cf. figures 1–3). The number of ERB scale according to Moore (1997a)¹ was adopted to reflect the perceptual significance of value changes. These ERBs are the equivalent rectangular bandwidths of non-rectangular filters that derive from the investigation of auditory frequency selectivity with the notch-noise method (Moore 1997b). It has a higher resolution than the Bark scale² for frequencies below 500 Hz, which concerns mainly F0 and F1. According to Traunmüller (1990), the concept of critical bands underlying the Bark scale is rather a measure of the tonotopic sensory scale, whereas ERB rate is a measure of frequency selectivity.

Functional Data Analysis: “(i) It takes account of the underlying continuity of the physiological system generating the behavior; (ii) it displays temporal dependencies in the data owing to this continuity; (iii) it provides methodologies to deal quantitatively with the complexities of multidimensional time series data like those collected in speech experiments ...” (Ramsay, Munhall, Gracco, and Ostry 1996)

3 The impact of position on personal pronouns

first person pronouns particularly frequent, opportunity to investigate repetitions of the same proclitic and enclitic sequences

¹ $E = 21.4 \log_{10}(4.37F + 1)$.

² $z = [26.81f / (1960 + f)] - 0.53$ (Traunmüller 1990).

3.1 /r/-vocalization in *wir*

- hypothesis: greater deviance from canonical form in enclitic position
- difficulty: following context not controlled
- *wir* frequent in enclitic position: 102 *können wir*, 22 *wir können*, 52 *haben wir*, 10 *wir haben*
- position in utterance. Simpson (1998): 233 ms for vowel in utterance-final *vor* vs 62 ms for non-final occurrences of the preposition. More time for diphthong in utterance-final position

In order to know how the r-diphthong in *wir* is represented in terms of formant movements, the formants were estimated automatically at five time points in the vowel and compared to the same values in monophthongal *wie*. The results are presented in table 1. It should be noted that since spontaneous speech is not controlled, standard deviations are much higher than in experimental setups where target items are embedded into a stable carrier sentence.

Figure 1 visualizes the data from table 1. Contrary to what one would expect for an opening diphthong, there are hardly any formant movements in *wir*, but the formants run almost in parallel with those for *wie* (for more discernibly converging formant movements in read *wir* cf. figure 7 on p. 14). However, there are three differences between the patterns for *wir* and *wie*: a) clearly lower F2 values for *wir*, b) clearly higher F1 values for *wir*, and c) a different shape of F1, which shows a slight upward movement for *wir*. This raise of F1 for *wir* seems to be a residuum of the opening gesture of the carefully pronounced r-diphthong. This holds for male as well as for female speakers. The phonetic correlates of the r-diphthong are thus not completely captured by the symbol [ɐ] which has often been used for ‘monophthongized’ productions.

To justify by an example the decision to plot perceptual ERB values rather than acoustic formant frequencies, figure 2 displays the values from table 1 without converting the Hertz values. One can easily see that the resolution is too high for high frequencies and too low for low frequencies. A logarithmic plotting of the frequency data, on the other hand, resolves low frequencies more than is perceptually relevant (figure 3).

Figure 4 displays the realization of the vowel in spontaneous speech depending on duration. The durational continuum has been divided into intervals of 30 ms (cases where the vowel was deleted are excluded, therefore the first class starts at 1 ms; the classes are displayed time-normalized, which

Table 1: F1 and F2 values and standard deviations for male and female speakers in unaccented *wir* and *wie*.

item	t	male						female							
		F1	sd	F2	sd	n	t_{obs}	sd	F1	sd	F2	sd	n	t_{obs}	sd
<i>wir</i>	0.1	396	59	1526	224	231	59	32	471	83	1753	276	155	58	23
	0.3	416	59	1529	211				493	76	1766	270			
	0.5	430	61	1514	204				504	73	1753	254			
	0.7	433	70	1487	209				505	89	1719	246			
	0.9	418	84	1467	226				482	109	1681	234			
<i>wie</i>	0.1	315	111	1984	171	41	54	13	322	46	2049	324	26	61	21
	0.3	305	87	2000	147				324	53	2019	316			
	0.5	305	84	2006	145				325	53	2013	325			
	0.7	308	108	1969	140				325	49	2019	229			
	0.9	325	200	1939	132				322	43	1939	177			

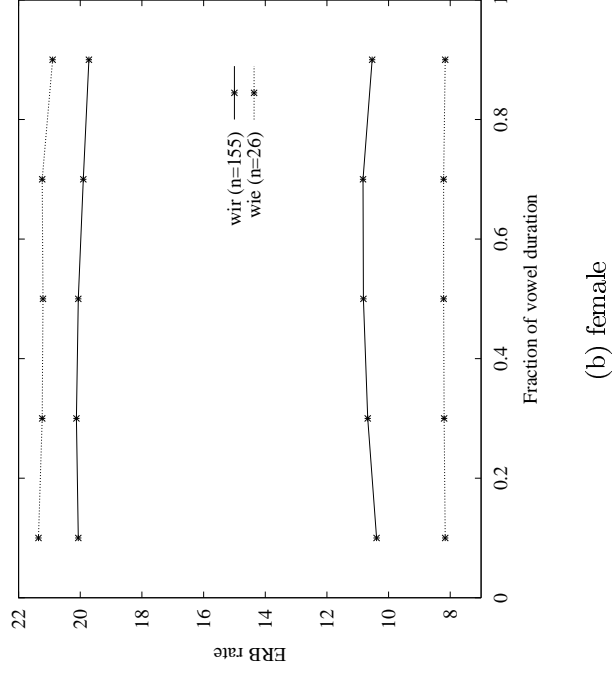
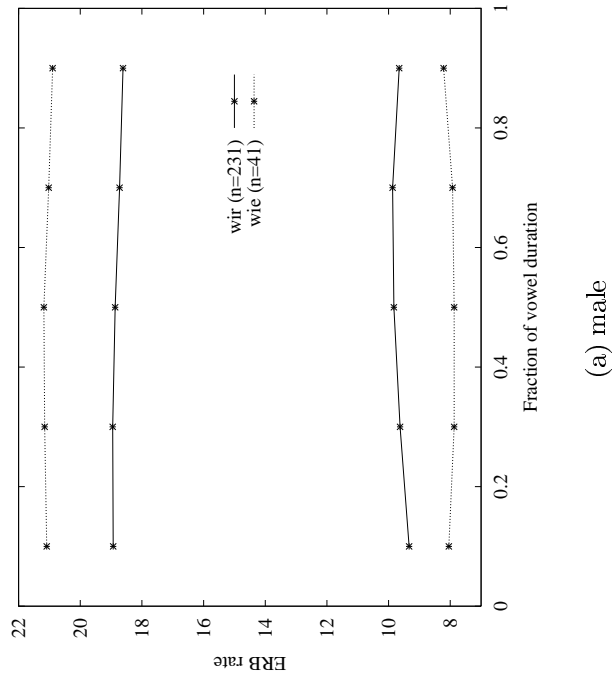
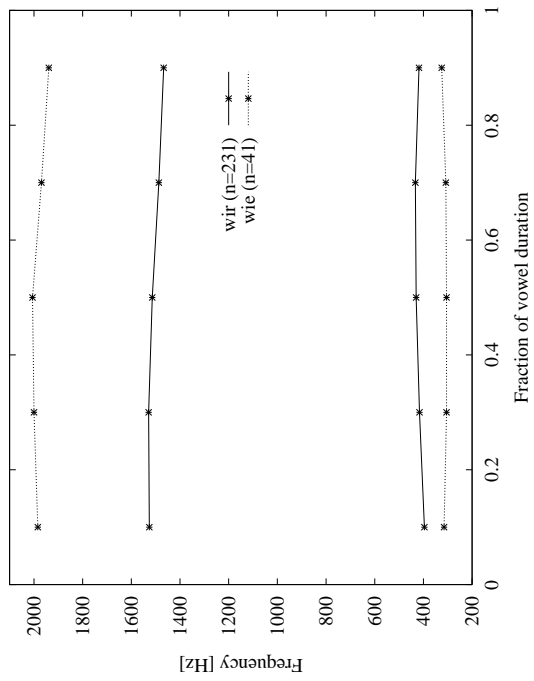
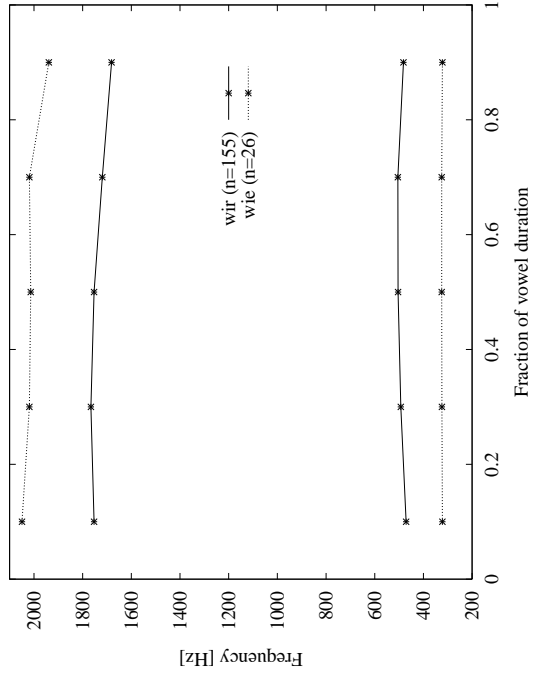


Figure 1: F1 and F2 for *wir* and *wie* in unaccented renditions without hesitational lengthening (prosodically labelled database), plotted in the perceptually relevant scale of ERB rate.

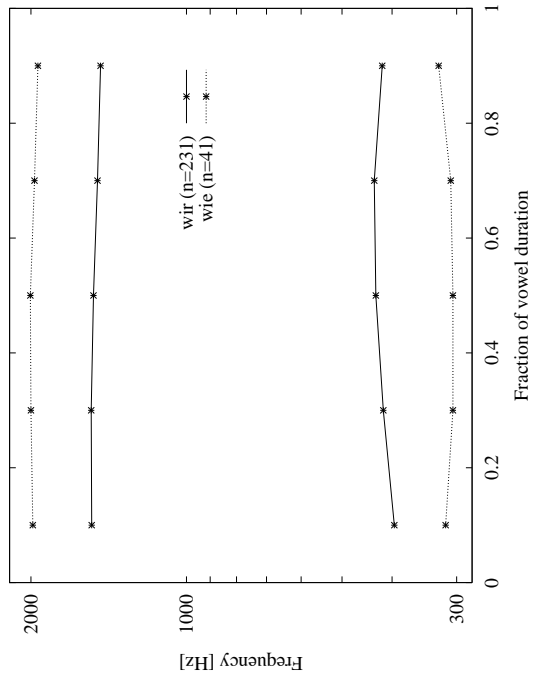


(a) male

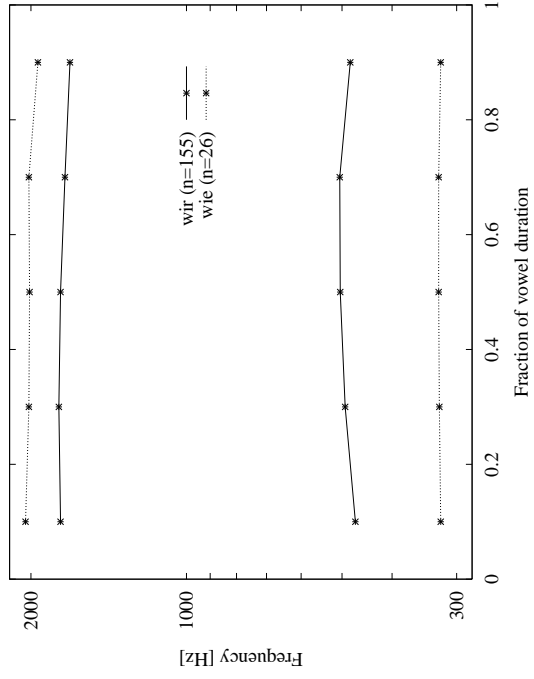


(b) female

Figure 2: F1 and F2 for *wir* and *wie* in unaccented renditions without hesitational lengthening (prosodically labelled database), the frequency data plotted linearly.



(a) male



(b) female

Figure 3: F1 and F2 for *wir* and *wie* in unaccented renditions without hesitations (prosodically labelled database), the raw frequency data plotted logarithmically.

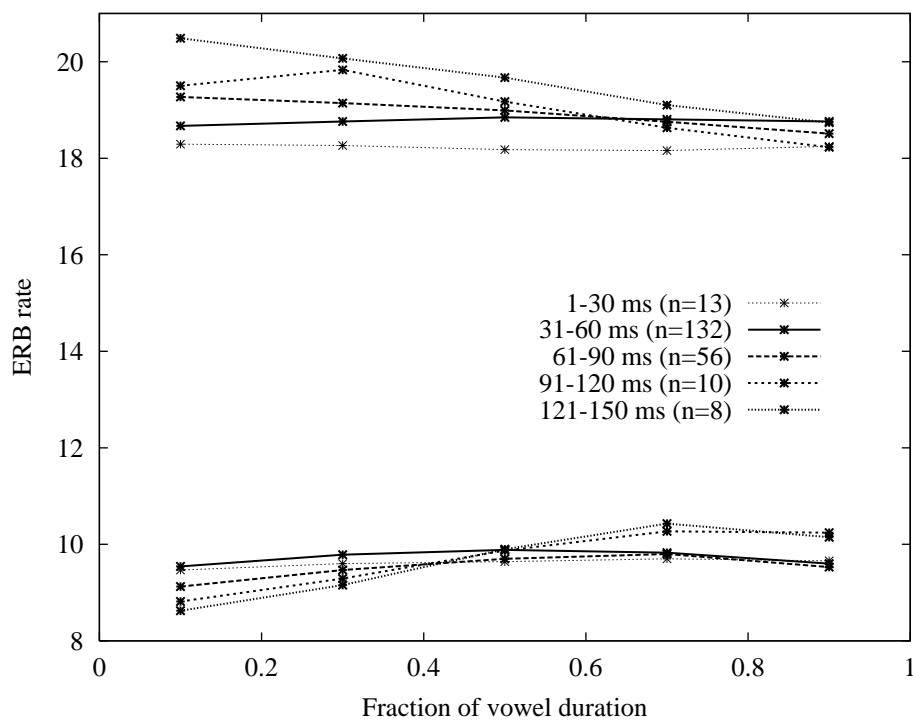


Figure 4: F1 and F2 in *wir* depending on duration (unaccented, male speakers).

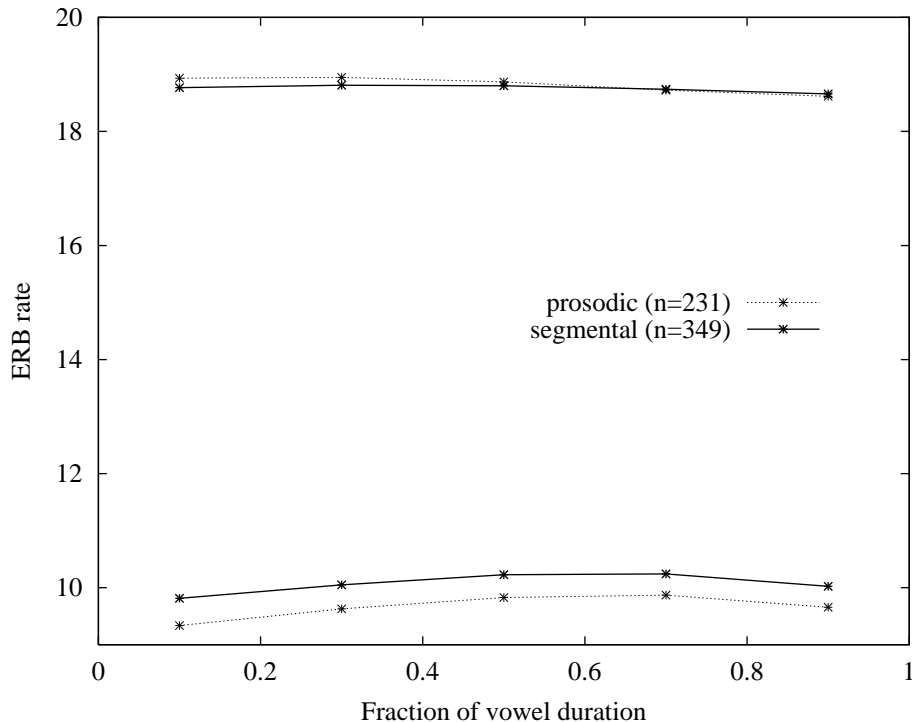


Figure 5: F1 and F2 of unstressed *wir* in prosodically and segmentally labelled databases (male speakers).

means that the inclinations of the graphs cannot be taken as measures for speed of acoustic change. For a pseudo-absolute representation of durational classes cf. figure 9). One can observe that with decreasing duration, the starting point of F2 is subsequently lowered, whereas the end-point remains almost constant, which results in a flattening of F2 at about the end-point level of the long production. F1, on the other hand, is rather flattened at the mid-point level of the long production.

After these preliminary investigations of the realization of *wir* in spontaneous speech in general, we now turn to the subject of reduction depending on position. Because the numbers of pre- and postverbal occurrences with the same auxiliary are extremely small when restricted to the prosodically labelled database, it makes sense to include tokens from the prosodically unlabelled database as well. The latter is missing reliable information on sentence-accent, i.e. some of the tokens which have not received a sentence-accent marker may be accented. Figure 5 displays the results for the two databases. The second formants are virtually identical, and the first for-

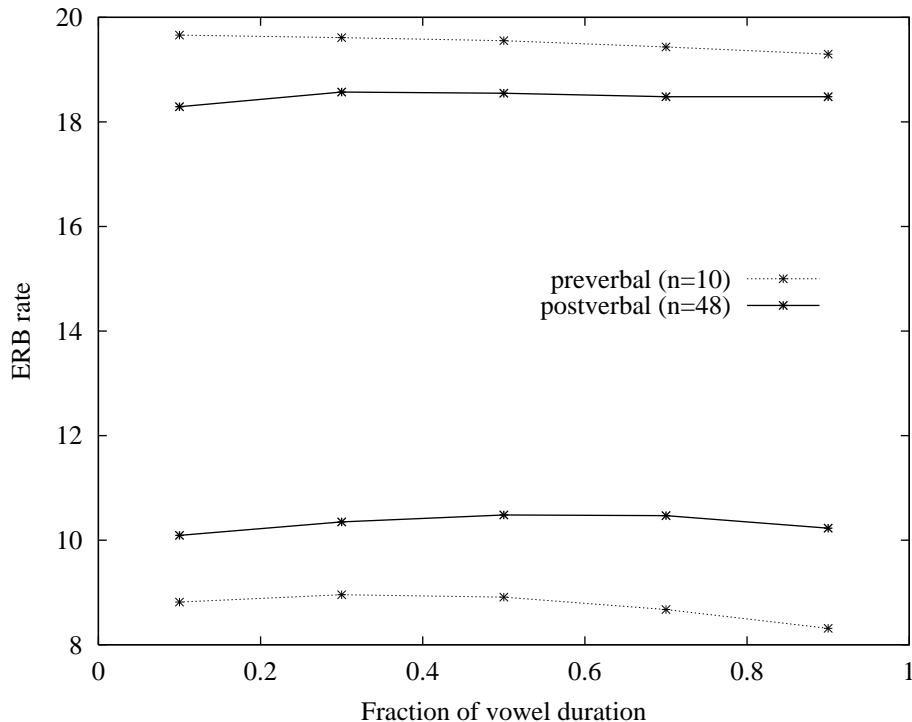


Figure 6: F1 and F2 of unstressed *wir* in *wir können* and in *können wir* (male speakers in complete database).

ments run in parallel, although F1 is slightly higher for the database which does not contain prosodic labels. The patterns found in the database without prosodic labels closely resemble the findings for the prosodically labelled database, and it seems legitimate to merge the databases for the investigation of positional factors.

Our expectation is that the deviance from the canonical form is greater in postverbal than in preverbal productions. Figure 6 shows that *wir* is more open when it follows *können* than when it precedes it. This is not due to increased duration in preverbal position, on the contrary, the preverbal vowels are even slightly shorter than the postverbal ones (means: 53 ms vs 61 ms). These results support our hypothesis of more postverbal reduction.

One difficulty in comparing the findings for pre- and postverbal vowels is that in the case of *wir können*, the immediate segmental context on both sides of the vowel is controlled, whereas in *können wir*, the following context is not controlled. The vowel is always followed by a velar plosive in *wir können*, but there are different postverbal contexts such as *können wir ja*

Table 2: Frequent sequences of *wir* and auxiliary verbs, n_{total} indicating the number of the sequence in the complete database without considering sentence-accent, and n_{prosun} indicating the number in the prosodically labelled corpus (both items unaccented). All numbers exclude cases with hesitational lengthening.

sequence	n_{total}	n_{prosun}
<i>wir können</i>	22	4
<i>können wir</i>	102	27
<i>wir haben</i>	10	2
<i>haben wir</i>	56	16
<i>wir wollen</i>	1	
<i>wollen wir</i>	32	1

and *können wir uns*. This means that part of the observed differences may be due to contextual factors.

- finer temporal resolution: measurements at 0.1, 0.3, 0.5, 0.7, and 0.9
- separated male and female speakers
table 1 and fig. 1
- computation of standard deviation within the script
- plotting data: changes in F1 smaller than in F2, not adequately captured in linear plotting of frequency values
 - first trial: logarithmic; changes exaggerated
 - Bark scale: low frequency resolution below 500 Hz; Traunmüller 1990: ‘CB should not be taken as a measure of frequency resolution, but CB rate may be taken as a measure of the tonotopic sensory scale.’ ($z = [26.81f/(1960 + f)] - 0.53$)
 - ERB rate as a measure of frequency resolution (Moore 1997b): writing and implementing function in awk ($E = 21.4\log_{10}(4.37F + 1)$)
- spontaneous vs read speech: formant movements bigger in read speech
fig. 7
- this also holds when separating durational classes
fig. 9 using ERB rate to plot frequencies

- so far only unstressed productions in prosodically labelled database, now data from segmental database for comparison

Because the numbers of pre- and postverbal occurrences with the same auxiliary are extremely small when restricted to the prosodically labelled database, it makes sense to include tokens from the prosodically unlabelled database as well. The latter is missing reliable information on sentence-accent, i.e. some of the tokens which have not received a sentence-accent marker may be accented. Figure 5 displays the results for the two databases. The second formants are virtually identical, and the first formants run in parallel, although F1 is slightly higher for the database which does not contain prosodic labels. The non-prosodic patterns closely resemble the prosodic ones, and it seems legitimate to merge the databases for the investigation of positional factors.

- all male productions of *wir* after *können* (spontaneous speech)

Figure 8 shows that *wir* is more centralized after *können* than on average of all its occurrences.

Figure 9 contrasts the realization of the vowel in spontaneous and read speech depending on duration. The durational continuum has been divided into intervals of 30 ms, and the three classes containing a sufficient number of production in both spontaneous and read speech are displayed. The five measurement points are projected onto the mean duration within the class, covering 80% of this duration (from 10% to 90%). The graphs for the three classes are aligned around their midpoints.

An important finding is that the quality of the vowel is more diphthongal in read speech, which shows clear upward movements of F1 and downward movements of F2 in all durational classes. The auditory and acoustic distance between the two formants is greater at the beginning and, in the case of the longest productions, smaller at the end. In spontaneous speech, the formants are flatter and converge to a smaller degree, or even slightly diverge in the case of the shortest productions.

Although these differences between the styles might be partly due to different class sizes and context distributions, there is a clear tendency in read speech to be more ‘conservative’, in the sense that it conserves the transition from a closed vowel to an open /r/-vocalization to a greater extent. In spontaneous speech, on the other hand, the /r/-vocalization manifests itself in a less linear way, with the starting and end points tending to level out each other.

It also turns out that reduction (in this case of formant movements) is not merely a consequence of temporal compression, since the durational

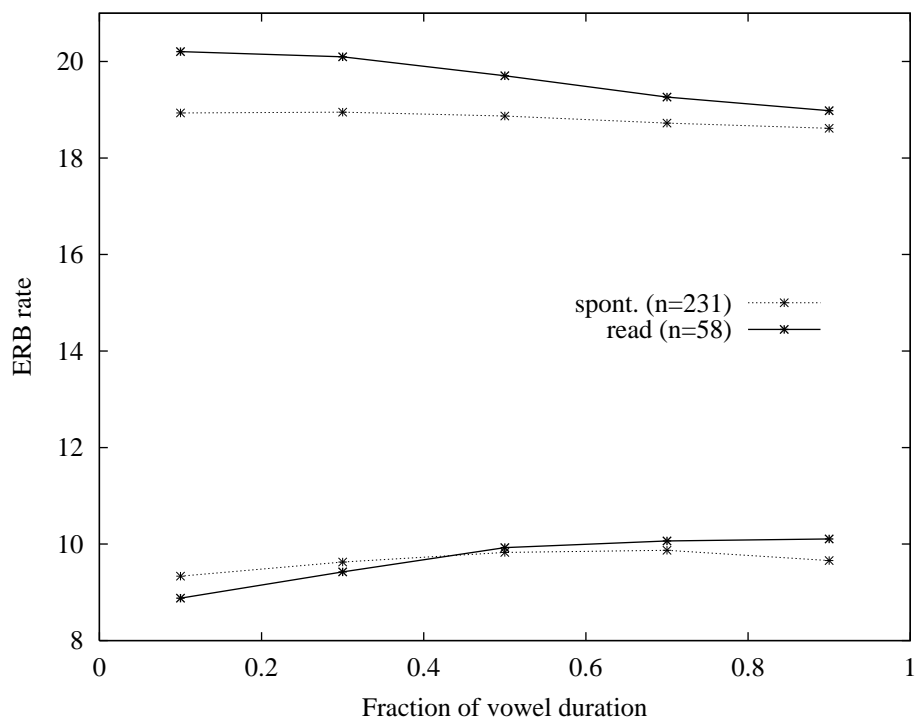
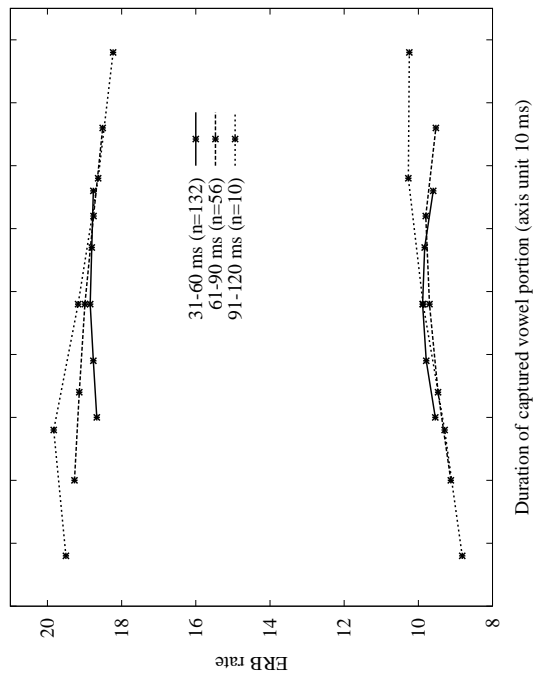
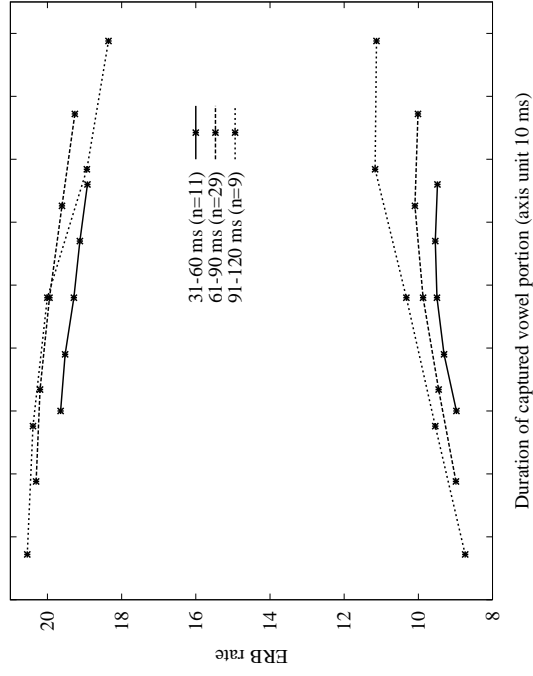


Figure 7: F1 and F2 for *wir* in read and spontaneous speech (male speakers, unaccented renditions without hesitational lengthening).

Figure 8: (temporary until problem with *wir können* is solved) F1 and F2 of unstressed *wir* in complete database and in *können wir* (male speakers).



(a) spontaneous



(b) read

Figure 9: F1 and F2 for three durational classes of *wir* in both read and spontaneous speech (unaccented renditions without hesitational lengthening). Frequencies converted into ERB scale for better representation of perceptual significance of formant movements.

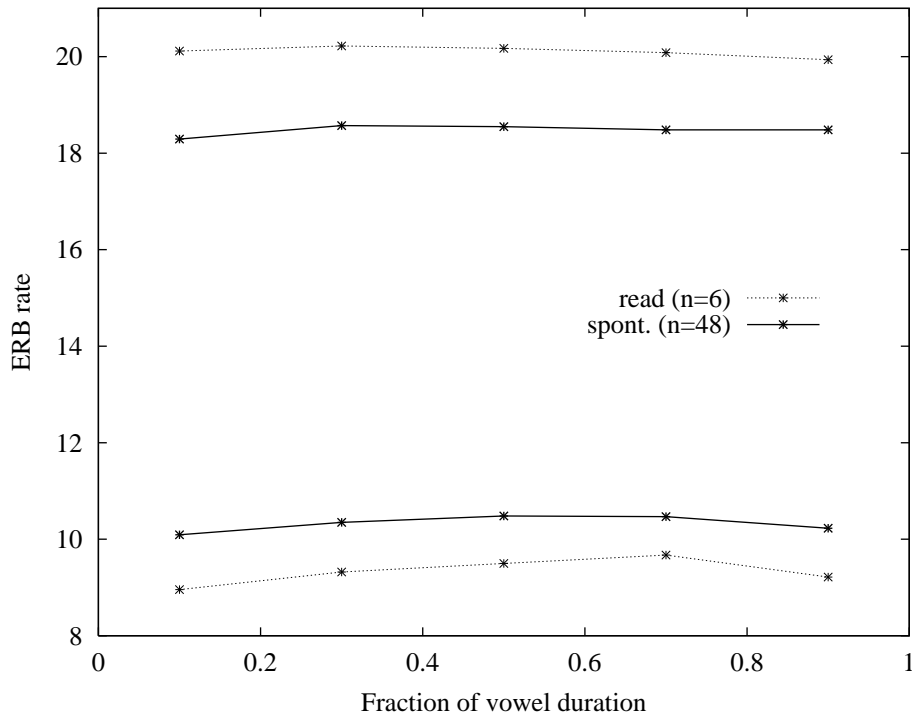


Figure 10: F1 and F2 of unstressed *wir* in *können wir* in read and spontaneous speech (male speakers).

parameter was controlled. The alignment of the speech movements is thus not completely left to the mechanics of the vocal tract who is forced to reduction by durational constraints; rather, the speaker selects a certain mode (stiffness?) according to style.

- influence of position in read speech cannot be investigated since *wir* occurs only in postverbal position: *Können wir nicht Tante Erna besuchen?* (be054)
- /r/-diphthong in *können wir* has a more peripheral and a slightly more diphthongal quality for F1 than in spontaneous speech (cf. figure 10)

3.2 Vowel shortening in *ich*

- sequence of palatal vocoid and contoid: shortening of vocoid
- hypothesis: shorter in proclitic position

Table 3: Frequent sequences of *ich* and auxiliary verbs, n_{total} indicating the number of the sequence in the complete database without considering sentence-accent, and n_{prosun} indicating the number in the prosodically labelled corpus (both items unaccented). All numbers exclude cases with hesitational lengthening.

sequence	n_{total}	n_{prosun}
<i>ich bin</i>	30	8
<i>bin ich</i>	128	57
<i>ich kann</i>	29	4
<i>kann ich</i>	98	16
<i>ich hab'</i>	45	4
<i>hab' ich</i>	105	31

Table 4: Mean segmental durations in *bin ich* ($n = 57$) and *ich bin* ($n = 8$).

segment	b(-h)	I	n	I	C
dur. [ms]	60	34	50	40	59

segment	I	C	b(-h)	I	n
dur. [ms]	48	68	83	32	75

- ideal for non-parametrical testing: 134 *bin ich*, 32 *ich bin*, 82 *kann ich*, 31 *ich kann*
- problem with *hab'*: vowel could also be exponent of schwa

The reduction of *ich* might take place in the durational domain, e.g. as the shortening of the vocoid in the sequence of palatal vocoid and contoid. Our hypothesis is that the vocoid is shorter when preceding the verb, since a devoiced onset seems to be more probable in phrase-initial sequences such as *ich kann* than in *kann ich* which mostly occurs phrase-internally. After discussing durational aspects, we turn to quality changes in the sequences.

Considering the combinations of *ich* and *bin* first, the tempo is higher in *bin ich* than in *ich bin*. Whereas the stretch corresponding to the first sequence lasts only 243 ms (144 ms + 99 ms), the latter sequence takes 306 ms (116 ms + 190 ms) (cf. table 4). The reason seems to be that *bin ich* is integrated into a stretch of verbal material, often exclusively function words, while *ich bin* occurs phrase-initially, and therefore is more exposed. The higher tempo for *bin ich* seems to be another reason for more reduction, besides the changes in context described above.

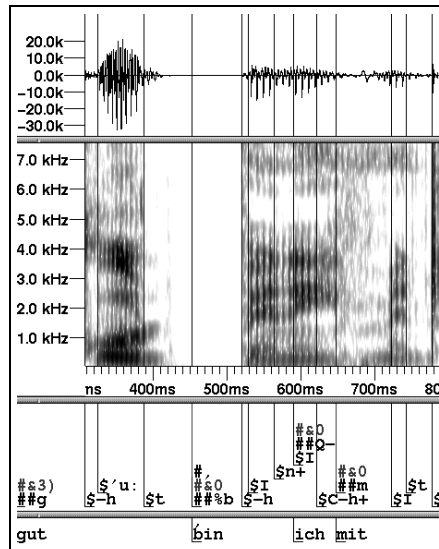


Figure 11: Postverbal reduction of palatal fricative in *bin ich mit* (g072a019).

Contrary to our expectation we see that the vocoid in *ich* is not shorter preverbally than postverbally. This does not only hold in absolute terms — which merely is a consequence of the slower tempo in the preverbal cases — but also in relative terms. The ratio of the vocoid divided by the contoid is 0.68 postverbally and 0.71 preverbally.

Turning now to quality changes, there are more reduction phenomena in *bin ich* than in *ich bin*. /b/ is less frequently released in *bin ich* (32 out of 57 vs 7 out of 8); this is because approximation is favoured in an intervocalic position such as *da bin ich*, but not after consonants as in *ich bin*. In no case of *bin ich* a glottal stop has been produced (2 cases in *ich bin*). The reason is that the vowel of *ich* occurs phrase-medially in *bin ich*, but phrase-initially in *ich bin* and mostly after non-verbal material like pauses or breath, which favour the presence of a glottal stop.

Figure 11 shows a reduction of the palatal fricative involving a vocoid with breathy voice, which seems to be restricted to sequences with a following nasal. For other replacements of dorsal gestures by laryngeal adjustments before nasals in the case of velar or uvular fricatives cf. Wesener (1999). It is obvious that this type of reduction can only occur postverbally, since the necessary nasal cannot be provided for preverbally.

The ratio of the vocoid divided by the contoid is 0.75 postverbally and 0.82 preverbally. Again, the preverbal vocoid is longer in both absolute and relative terms, for the same reasons as discussed above in connection with

Table 5: Mean segmental durations in *bin ich* ($n = 16$) and *ich kann* ($n = 4$).

segment	k -h	a	n	I	C
dur. [ms]	105	53	51	45	60

segment	I	C	k -h	a	n
dur. [ms]	51	62	84	54	29

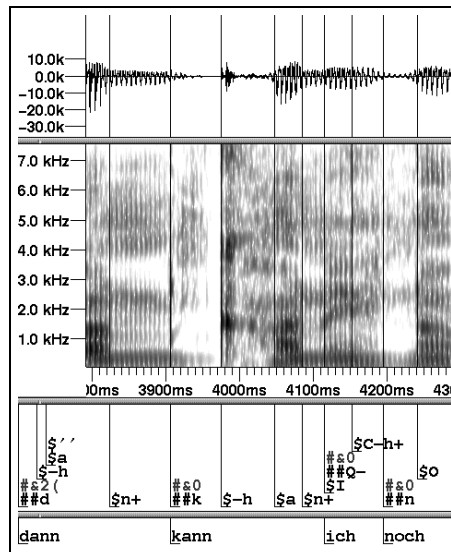


Figure 12: Postverbal reduction of palatal fricative in *kann ich noch* (g315a007).

bin sequences.

Replacement of the dorsal by a laryngeal gesture (breathy voice) before nasals is illustrated in figure 12 and restricted to postverbal items (see above).

problem with *hab'*: vowel could also be exponent of schwa (blending of vowel qualities after dropping of glottal activity)

- basis: complete spontaneous database (with and without prosodic labels)
- build lists containing all verbs that can follow or precede *ich*
- extract relevant lines from a lexicon with two-word entries, using named arrays in *awk*

- preverbal: 541 sequences with 60 verb types (the latter not counting forms with apostrophe)
- postverbal: 726 sequences with 50 verb types (the latter not counting forms with apostrophe)
- vowel in preverbal *ich* more frequently ‘deleted’ (frequently correlates of the vowel are present, **ma**):
 - preverbal 40 (7%, 15 with **ma**)
 - postverbal 20 (3%, 8 with **ma**)
- more often absence of glottal reflex in postverbal *ich*:
 - preverbal (501): 178 Q- (36%)
 - postverbal (706): 542 Q- (77%)
- no difference in consonant ‘deletion’:
 - preverbal 11 (2%, 7 with **ma**)
 - postverbal 17 (2%, 9 with **ma**)
- consonant ‘replacement’: glottal fricative only in postverbal *ich*
 - preverbal 6 (1%): labiodental, alveolar and postalveolar fricatives
 - postverbal 10 (1%): labiodental, alveolar and glottal fricatives
- extracting all cases of *ich meine* from the lexicon
- checking ambiguous cases in the transliteration files
- *meine* wrongly marked as function word: g086a010, g364a007, g372a014, g375a003; corrected
- preverbal: 454 with 20 verb types (the latter including forms without final <e>; trap: missing upper case *Ich*)
- postverbal: 195 with 17 verb types (the including forms without final <e>)
- vowel in preverbal *ich* more frequently ‘deleted’:
 - preverbal 62 (14%)
 - postverbal 7 (4%)

- more often absence of glottal reflex in postverbal *ich*:
 - preverbal (392): 126 Q- (32%)
 - postverbal (188): 160 Q- (85%)
- consonant ‘deletion’
 - preverbal 2 (0%)
 - postverbal 2 (1%)
- consonant ‘replacement’
 - preverbal 3 (1%)
 - postverbal 1 (1%)

In both read and spontaneous speech, the vowel is more often deleted in preverbal than in postverbal *ich*. This finding corroborates an earlier observation which was not yet based on large computer corpora of speech (Kohler 1979). The vocoid and contoid in *ich* can be produced with the same tongue configuration: devoicing leads to a stronger airstream, which then causes friction at the palatum. When preverbal *ich* occurs utterance-initially (which is not possible postverbally), it is easier to maintain the open position of the vocal folds associated with breathing throughout *ich*, which results in a contoid only.

- spontaneous speech: cf. table 6; ‘other’ comprises all contexts that do not contain a verb in the first person singular
- total number in complete database greater than in Wesener (1999): 1410 instead of 1360. This is mainly because the present corpus contains five more dialogues. Apart from that, the numbers also differ slightly because the present numbers derive from search operations on the orthographic field of a lexicon only, whereas the earlier numbers were calculated using Kieldat functions on different fields of the database.
- read speech: cf. table 7

Table 6: Position of *ich* in the spontaneous database.

context	n
preverbal	541
postverbal	726
other	143
total	1410

Table 7: Position of *ich* in the read database.

context	n
preverbal	454
postverbal	195
other	72
total	721

3.3 Centralization in *Sie*

- hypothesis: more central in enclitic position (Kohler 1979)
- difficulty: following context not controlled
- 38 *haben Sie*, 6 *Sie haben*, 15 *können Sie*, 4 *Sie können*
- influence of sentence-accent with *können*
- hypothesis: more central postverbally (Kohler 1979)
- difficulty: following context not controlled

Table 8: Frequent sequences of *Sie* and auxiliary verbs, n_{total} indicating the number of the sequence in the complete database without considering sentence-accent, and $n_{pros\ un}$ indicating the number in the prosodically labelled corpus (both items unaccented). All numbers exclude cases with hesitational lengthening.

sequence	n_{total}	$n_{pros\ un}$
<i>Sie haben</i>	5	2
<i>haben Sie</i>	36	11
<i>Sie können</i>	5	-
<i>können Sie</i>	15	4

- influence of sentence-accent with *können*

In a first step, all verbs that preceded or followed *Sie* were extracted from the context lexicon. Items that occurred as infinitive and imperative plural were excluded, as was a slip of the tongue. Only sequences of *Sie* as a subject pronoun plus verb in the third person plural were thus examined.

Again, the item was more frequent postverbally, where it also occurred with a greater variety of verb types (postverbally 33, preverbally 13 different types). Figure 13 displays formant movements in pre- and postverbal position for male and female speakers in spontaneous speech.

In accordance with our expectation, male speakers show a lower F2 postverbally, which points to a less anterior production. Female speakers, however, seem to make no difference in F2 between the positions.

For both genders, F1 moves upward towards the end of the vowel in preverbal position. Clearer formant transitions may be a sign of prominence. The rise, however, may also be an artefact caused by the comparatively low number of tokens in preverbal position.

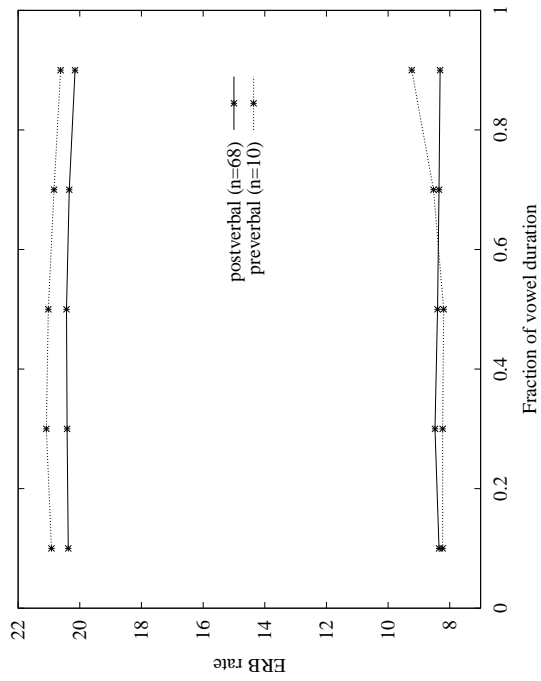
Figure 14 shows the results for read speech. In accordance with our expectation, the postverbal formants lie within the preverbal formants for both male and female speakers³.

Whereas F1 clearly differs only in the last phase of the vowels in spontaneous speech, it differs throughout the vowel in read speech. F2 for female speakers, which is not different in spontaneous speech, is differentiated across the positions in read speech. Although the samples are small, there is thus a tendency towards a clearer distinction of position in read speech.

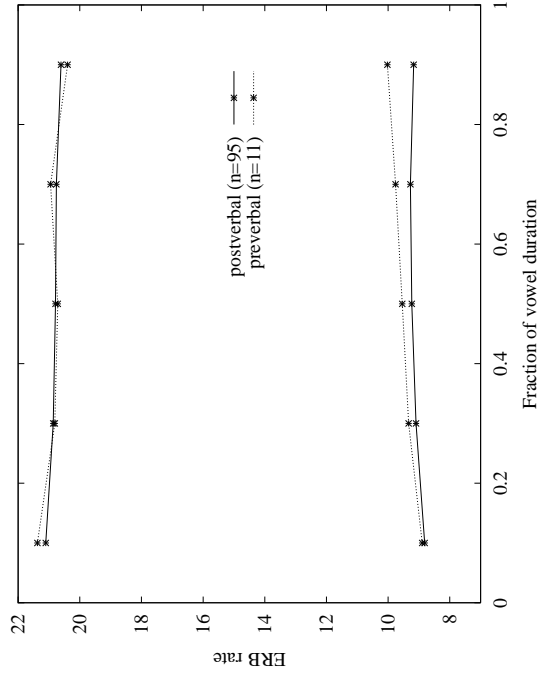
3.4 *es* and *'s*

- new lexicon without calling function `cvApostrophe` to retrieve only *es* and not completed *'s*
- basis: complete spontaneous database (with and without prosodic labels)
- lists containing all verbs that can follow or precede *es* (only third person singular)

³Problems in the automatic calculation of F2 occurred for the preverbal productions of female *Sie*: occasionally, an additional resonance was wrongly placed in between F1 and F2. This seems to be due to the blurred spectral characteristics of the vowel in the vicinity of fricatives, such as in *Sie wollen*. Additional resonances were not removed in the data files in order to keep the results compatible with those for postverbal position, which were not changed by hand either.

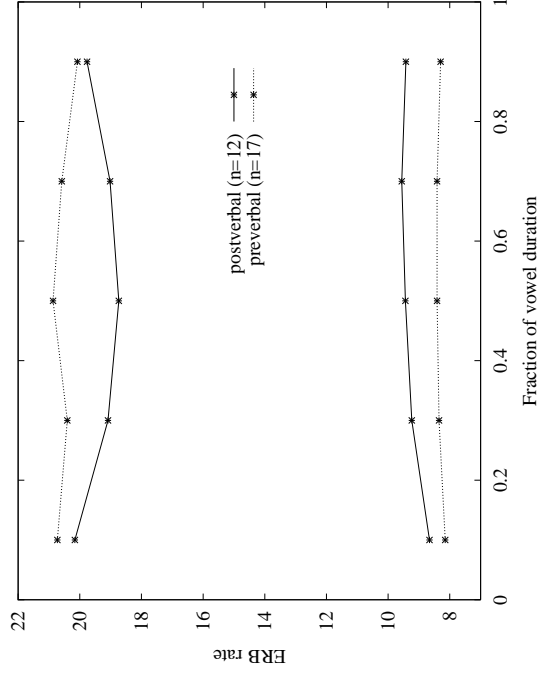


(a) male

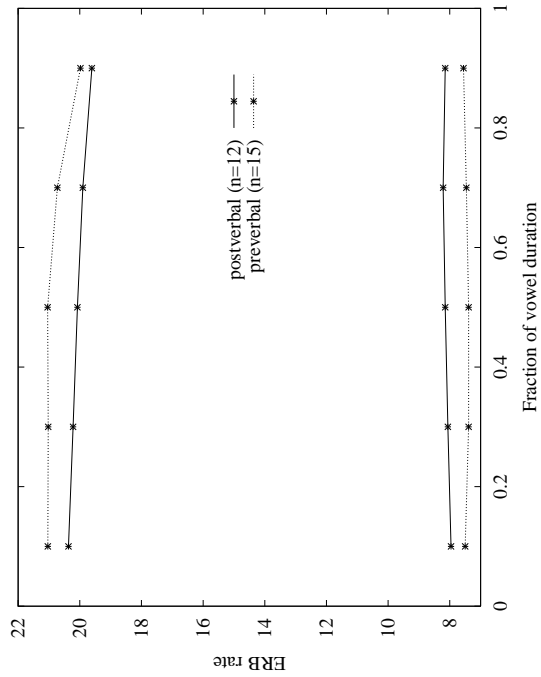


(b) female

Figure 13: F1 and F2 for unaccented *Sie* in pre- and postverbal position (complete spontaneous database).



(a) male



(b) female

Figure 14: F1 and F2 for unaccented *Sie* in pre- and postverbal position (complete read database).

Table 9: Position of *es* in the spontaneous database.

context	<i>n</i>
preverbal	52
postverbal	182
other	53
total	287

- preverbal: 52 sequences with 21 verb types (the latter including forms with apostrophe)
- postverbal: 182 sequences with 24 verb types (the latter including forms with apostrophe)
- vowel in *es* deleted:
 - preverbal 6 (12%)
 - postverbal 17 (9%)
- more frequently absence of glottal reflex in postverbal *es*:
 - preverbal (46): 18 Q- (39%)
 - postverbal (165): 123 Q- (75%)
- pre- and postverbal *'s* in spontaneous speech:
- preverbal: 9 sequences with 6 verb types (third person singular, including forms with apostrophe)
- postverbal: 95 sequences with 12 verb types (third person singular, including forms with apostrophe)
- vowel in *'s* deleted:
 - preverbal 8 (89%)
 - postverbal 93 (98%)
- *es* and *'s* in spontaneous speech:
- although a few productions of *'s* are ambiguous in that they could also derive from *das*, it is legitimate to combine the two orthographic items

Table 10: Position of 's in the spontaneous database.

context	<i>n</i>
preverbal	9
postverbal	95
other	59
total	163

Table 11: Position of *es*/'s in the spontaneous database.

context	<i>n</i>
preverbal	61
postverbal	277
other	112
total	450

- vowel in *es*/'s deleted:
 - preverbal 14 (23%)
 - postverbal 110 (40%)
- the pronoun occurs more frequently in postverbal position
- vowel is more frequently absent in this position
- vowel misses in connection with a subset of frequent verbs that are mostly function words themselves
 - preverbally: *besteht* (only once) *ginge ist müßte tut wär' war wird*
 - postverbally: *geht gibt ginge ist hab' paßt sah' schaut sieht wär'/wäre war wird würd'/würde*
- Pre- and postverbal *es* in read speech:
- only verbs in third person singular
- preverbal: 62 sequences with 6 verb types, all sentence-initial
- postverbal: 131 sequences with 11 verb types
- vowel in *es* deleted:

Table 12: Position of *es* in the read database.

context	<i>n</i>
preverbal	62
postverbal	131
other	2
total	195

- 0 preverbal (%)
- 10 postverbal (%)
- more frequently absence of glottal reflex in postverbal *es*:
 - preverbal (62): 9 Q- (15%) (most frequently glottal stop plus creak, 44%)
 - postverbal (121): 104 Q- (86%)

In both read and spontaneous speech, lack of glottal activity is much more frequent in postverbal vs preverbal *es*. The difference between the positions is even greater in read speech than in spontaneous speech since all preverbal tokens of *es* in read speech occur utterance-initially, thus favouring glottal activity.

As to 's, there were 94 tokens in the read database, all labelled with an alveolar fricative only. It makes little sense to compare this number with the findings for spontaneous speech with regard to the frequency of vowel elision. In spontaneous speech, transliteration followed speech, whereas in read speech, orthography preceded speech, thus more or less prescribing when a vowel should or should not occur.

3.5 Outlook on new scenario: *er*

- new speakers incorporated into KielDatGender
- in more than half of the turns (60 out of 116) alignment errors
- these errors probably result from overlaid #
- resulting database not yet reliable enough to measure formants in *er*

4 Other phenomena in read speech

4.1 /x/

- 2.8% (out of 1012 occurrences of /x/) x- in *noch*, *auch*, *doch*, *nach*; 2.9% x-h, frequently in *nach*; in sum 5.7%, which is significantly less than 17% in spontaneous speech. This might in part be due to the more frequent use of changed labels in spontaneous speech which was segmented later than the read corpus
- dorsal articulation can be given up in favour of glottal activity: *noch sagen* [nɔ̃fizaʊŋ] (dlme054, cf. figure 16), *auch noch* [aʊ̃finɔ̃χ] (hpte052), similarly *nach Mannheim* (kkoe062)
- breathy voice for /x/ can also be shifted, in one case the correlate of /x/ seems to be produced in the vowel of the following word: *noch die Verbindung* [nɔ̃dɪf-] (dlme010, cf. figure 15); breathiness does not seem to be merely induced by the following voiceless fricative because it is stronger at the beginning of the vowel than immediately before the fricative
- *noch andere* (kkoe061): /x/ marked by drop in F1 and F2 as well as rise in F4 (cf. figure 17 and the sections in figure 18); production may partially be the correlate of the glottal activity connected with the following vowel
- the interaction between vowel-related glottalization and glottal activity for /x/ can even be more complex. In *auch eine* (cf. figure 19, ugae073), the end of the diphthong in *auch* is produced with creaky voice. This creaky voice seems to be too strong to simply announce the short glottal stop at the beginning of *eine*. The creak has been labelled as being related to the vowel of *auch*. It may be that shifting the creak to the right fulfils two functions, i.e. boundary marking in connection with the vowel of *auch*, and indicating /x/. In addition, there may be a velar approximation around the stop.
- sometimes the dorsal fricative is present, but the borderline to the following labiodental fricative is not easy to detect: *noch von* (dlms068), *auch für* (kkoe062); one reason is that both fricatives have comparatively little intensity, and in addition the independent labial and dorsal articulations probably overlap

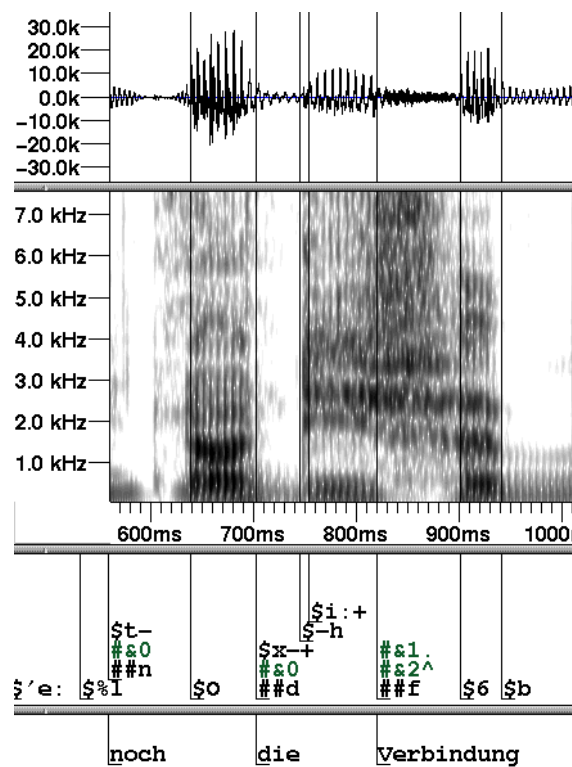


Figure 15: Correlates of /x/ in *noch die* (dlme010).

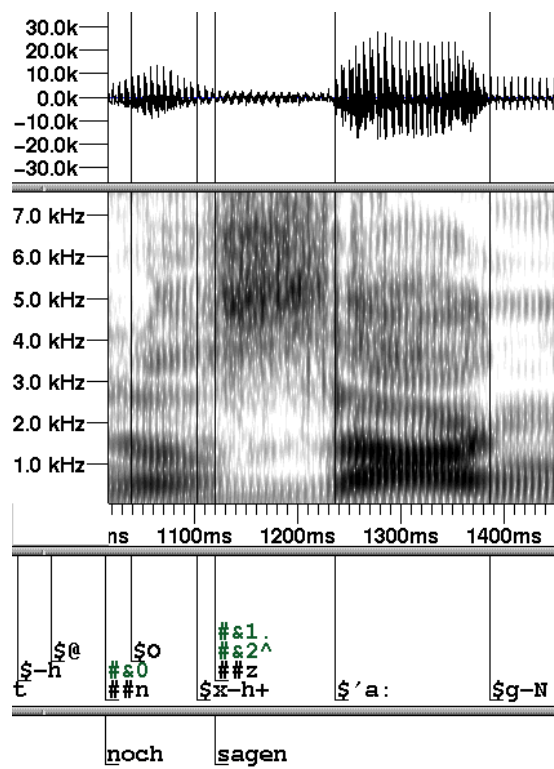


Figure 16: Correlates of /x/ in *noch sagen* (dlme054).

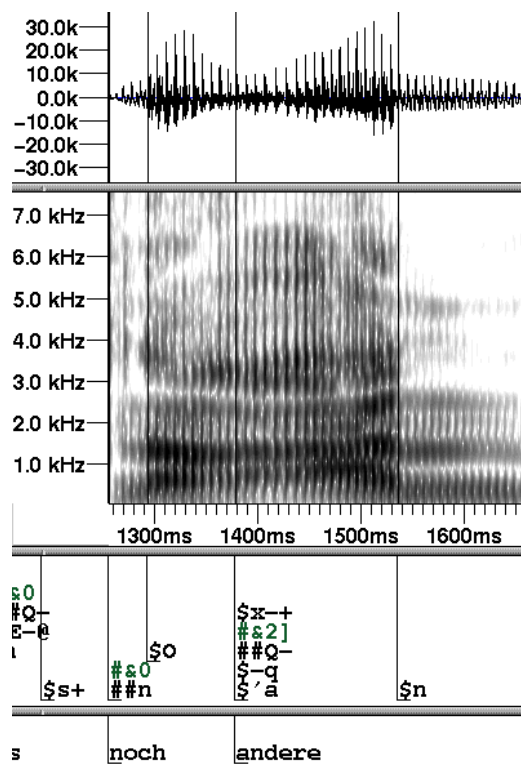


Figure 17: Correlates of /x/ in *noch andere* (kkoe061).

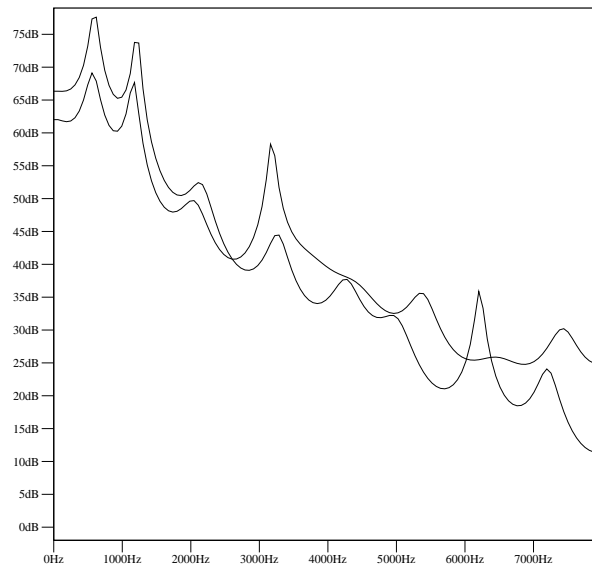


Figure 18: Sections in the vowel of *noch* (high-energy F1 and F2) and in the stretch corresponding to /x/ (high-energy F4, *kkoe061*).

- this overlap can even become stronger. In one case of *auch schon* the postalveolar fricative has a formant around 1200 Hz, and its energy cut-off lies as low as 700 Hz. Its two surrounding segments, i.e. the fricative in *auch* and the vowel in *schon*, have been marked deleted. The fricative is strongly rounded, and the rounding serves as the correlate of the vowel in *schon* (marked by *-ma*) which is not present in a linear sense. The fricative is velarized more than would be necessary for the vowel of *schon*, there might even be a primary postalveolar-velar double articulation. In addition, the fricative is preceded by a breathy-voiced stretch which together with the velar component of the fricative seems to express /x/ (*k22butt2*, cf. figure 20; in contrast to the original text, *als* has been produced after *schon* and not before).
- one case where *x* has been deleted in *doch* displays short dorsal friction in the transition to the velar stop in *doch gleich* (*k69mr092*)
- cross-language perspective: the close connection between [f] and dorsal fricatives can also be found in Czech: in voicing assimilation, the two behave like a voiced/voiceless pair (Dankovičová 1999).
- in a diachronic perspective, Czech has lexicalized the production of [f] for velar articulations that occur in the cognates of other Slavic

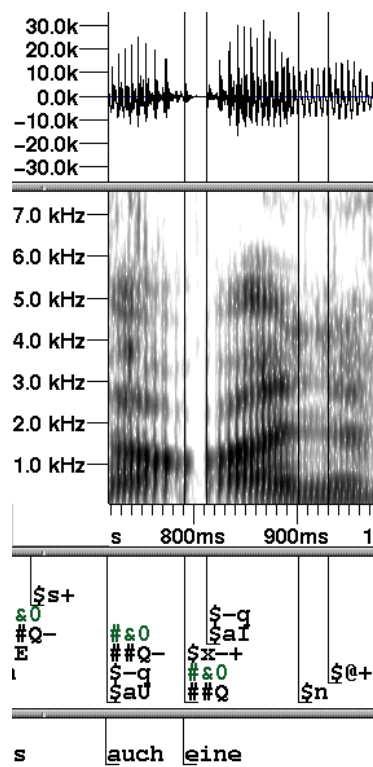


Figure 19: Creaky voice in *auch eine* (ugae073).

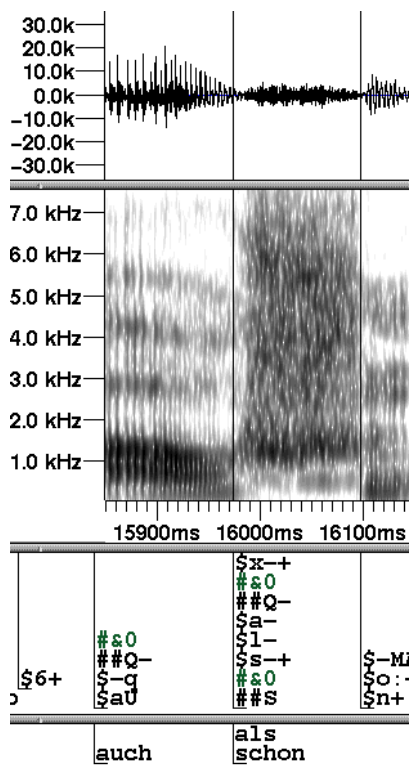


Figure 20: Breathy voice and the velar component of the following postalveolar fricative as correlates of /x/ in *auch schon* (k22butt2).

languages (Olle Engstrand, personal communication)

- Standard Chinese: the correlates of what is symbolized with <h> in the official *pinyin* system of transliteration vary, it can be a voiceless velar (Ladefoged and Maddieson 1996) or uvular fricative (Norman 1988). Often there is little frication even in careful speech, it seems probable that merely glottal articulations can be found in connected speech
- in all these cases glottal adjustments cause a blurring of the formant structure which loosely resembles the spectral characteristics of back dorsal articulations
- speaking styles: contrary to what one might expect, read speech can display quite radical deviations from standard dorsal fricatives, although these forms seem to occur less frequently than in spontaneous speech

4.2 /h/ in read speech

h has been marked deleted in only 1.8% of the cases, as opposed to 9.2% in spontaneous speech. This points towards more care with regard to glottal activity and its delimiting function, probably induced by orthography.

The item *Bahnhof*, however, is frequently produced without glottal activity delimiting the second morph (5 out of 24 tokens). The compound noun means ‘train station’, and the meaning of the first element is indeed ‘train’. But from a synchronous perspective, the second element *-hof* ‘courtyard; farm’ is opaque since a train station neither resembles a courtyard nor a farm. The reduction of glottal activity seems to indicate that *Bahnhof* is interpreted as mono- rather than bimorphemic. This tendency is probably even stronger in spontaneous speech, where the speaker is not influenced by orthography; unfortunately, the item does not occur in the present corpus of spontaneous speech.

How is word-initial /h/ realized after vowels in read speech? Unfortunately, the sequence *da haben* does not occur in read speech. A comparatively frequent sequence involving /h/ in intervocalic position is *Vater hat* in *Vater hat den Tisch gedeckt* ‘Father has set the table’. Most cases involve sequential productions with a medial portion of breathy voice (k02, k03, k05, k06, k61, k62, k66). A medial decrease in amplitude with the second half of the vocoid portion more breathy than the first half is found twice (k04, k64). In one production, breathy voice is shifted to the end (k63), and one token is breathy throughout (k65). Read speech thus seems to show similar temporal flexibility of the glottal correlates of /h/ as spontaneous speech.

4.3 Vowel nasalization

In read *uns*, the nasal contour is always present ($n = 64$). The amount of productions without contour for *und*, however, is very similar to the one found in spontaneous speech (1%, $n = 551$). For read *uns*, the orthographic presence of <n> seems to block the tendency for realizations without nasal contours, which in spontaneous speech is much stronger for *uns* than for *und*. The result may be interpreted as supporting the hypothesis that orthography has a conserving effect on pronunciation.

5 Discussion

- lexicalization/cliticization, also in comparison to other languages (e.g. Abercrombie 1964)
- relation spontaneous/read speech
- initial lengthening in parallel to final lengthening to mark adjacency to phrase boundary?

Work with corpora of connected speech increases the knowledge of phenomena connected with certain phonological units, such as /x/, /h/, and /r/ in this paper. But the impact of these findings reaches beyond the level of description, it has consequences for models of the phonology-phonetics interface and of phonology which are set up to account for the phonetic data.

Componential approaches seem promising when it comes to accounting for the phonetic data reported here. Articulatory Phonology (AP, Browman and Goldstein 1992) can adduce the two mechanisms of gestural overlap and reduced magnitude to account for the observations in connection with /x/ and /h/. But AP cannot explain ‘long’ non-sequential productions by only referring to increased overlap since overlap, other variables remaining constant, leads to shorter durations. It seems that AP must postulate a changed stiffness in these cases. This mechanical explanation, however, does not seem to capture the functional use speakers make of the mechanics of their vocal tracts.

The Window Model of Coarticulation (WM, Keating 1990) takes sequential allophonic feature specifications as its input and selects the possible range of physical values, i.e. the window, corresponding to each segmental feature. The behaviour of a given articulator during an utterance is then modelled by the interpolation across the windows of neighbouring segments. The model is devised for contextual phenomena in a linear sequence of segments: “Coarticulation refers to articulatory overlap between neighboring segments, which

results in segments generally appearing assimilated to their contexts” (Keating 1990, p. 452). A window is attributed to each segmental slot, and the temporal alignment of windows seems to be fixed. It is not clear how the model can deal with productions that display all exponents of different phonological units simultaneously, and that do not show a sequential organization of phonetic correlates any longer.

Whereas AP and WM are models of the phonology-phonetics interface getting their input in the form of sequential phonological specifications, Prosodic Analysis (PA) emphasizes the fundamental phonological relevance of phenomena beyond sequences of contrastive sounds derived at the word level. PA’s concept of prosodies seems suited for capturing non-sequential aspects of speech. Firth (1948) regards vowel-initial glottalization in German as a prosody marking the junction between lexical items. In a similar vein, he interprets breathiness in connection with English /h/ as a prosodic signal of initiality.

Firth observes that “the aitchiness, aitchification, or breathiness of sounds and syllables, and similarly their creakiness or ‘glottalization’ are more often than not features of the whole syllable or set of syllables” (Firth 1948, p. 146). In the investigated corpus, productions are indeed found where non-modal glottal activity is not limited to one segment or to a segment boundary, but spreads over longer stretches (e.g. *auch noch mal* produced as [ɑ̥f̥iŋəf̥im-], g415a004); these ‘global’ productions, however, do not seem to outnumber more limited stretches of glottal activity in German. The central aspect of Firth’s observation is that glottal activity is not localized at a certain point within sequences of phonological units, and this also captures the German examples with limited glottal activity that is shifted within or across syllables.

Extrapolating the prosodic interpretation to word-final dorsal fricatives in German, one might regard them as junction prosodies signalling word finality. In cases where their only exponent is non-modal glottal activity, this activity shows the same temporal flexibility as the other glottal prosodies.

Why are non-modal phonation types prominent candidates for non-sequential productions? One reason probably is that laryngeal activity is comparatively independent from supralaryngeal articulations. But this independence, instead of being left to the mechanics of the vocal tract, is in many cases put to a perceptual function. Articulatory interpretations do not imply a mechanical necessity of certain forms, they only state why it is plausible that such forms occur. Articulation is influenced by mechanical constraints, but it is mainly determined by communicative and social factors that prescribe, allow or prohibit certain phenomena.

Procedural interpretations that derive ‘reduced’ productions from underlying phonological forms via a set of rules bear the danger of missing important aspects of the investigated material (cf. Simpson 1992). From a procedural point of view, one might postulate that the vowel preceding /r/ is ‘elided’ in monophthongal productions of *wir*: [ɪɐ] → [ɐ]. A declarative understanding of the relation between the phonological units and their phonetic correlates seems more appropriate: both phonological units are phonetically coded in the monophthong.

The vowel-/r/ combinations are often said to involve /r/-‘vocalization’, again implying a procedural derivation. In inalterable items like the pronoun *wir*, there is no paradigm with an alternation between vocoid and contoid correlates of /r/. Rather than being a ‘vocalized’ realization of consonantal /r/, the vocoid broadly transcribed as [ɐ] is the exponent of /r/, the latter being an abstract phonological unit and neither a contoid nor a vocoid. In this context, /r/ is not fixed sequentially to the end of the vocalic segment and may be interpreted as a prosody.

A common belief is that ‘reduction’ is the result of increased speech rate. Since unscripted speech is supposed to be faster than read speech, it is expected to display more radical deviations from isolated word forms. The findings for /r/, however, show that it is important to disentangle style and speech rate. Similar durational patterns across speaking styles can be connected with different productions. In these cases, the phonetic coding is determined by style and not by duration, i.e. by communicative function and not by the mechanics of the vocal tract.

In a cross-language perspective, temporal flexibility of glottal correlates for /x/ and /h/ is probably not exclusive to German. The variation of dorsal fricatives with breathy voice and breath in German is found as an alternation in Czech voicing assimilation (Dankovičová 1999), and it seems worthwhile to investigate whether non-sequential productions of the glottal correlates are found in Czech. /h/-correlates in English and Swedish also seem promising areas to extend our knowledge on non-sequential aspects of speech.

References

- Abercrombie, D. (1964). Syllable quantity and enclitics in English. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. M. Trim (Eds.), *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961*, pp. 216–222. London: Longmans.
- Browman, C. P. and L. Goldstein (1992). Articulatory phonology: An

- overview. *Phonetica* 49, 155–180.
- Dankovičová, J. (1999). Illustrations of the IPA: Czech. In IPA (Ed.), *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, pp. 70–73. Cambridge: Cambridge University Press.
- Delsarte, P. and Y. V. Genin (1986). The Split Levinson Algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34(3), 470–478.
- Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*, 127–152.
- IPDS (1994). *The Kiel Corpus of Read Speech*, Volume 1, CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1995). *The Kiel Corpus of Spontaneous Speech*, Volume 1, CD-ROM#2. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1996). *The Kiel Corpus of Spontaneous Speech*, Volume 2, CD-ROM#3. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1997). *The Kiel Corpus of Spontaneous Speech*, Volume 3, CD-ROM#4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. In J. Kingston and M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pp. 451–470. Cambridge: Cambridge University Press.
- Kohler, K. J. (1979). Kommunikative Aspekte satzphonetischer Prozesse im Deutschen. In H. Vater (Ed.), *Phonologische Probleme des Deutschen*, Number 10 in Studien zur deutschen Grammatik, pp. 13–39. Tübingen: Narr.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling*, pp. 69–92. Dordrecht: Kluwer.
- Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Matthews, P. H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford and New York: Oxford University Press.
- Moore, B. C. J. (1997a). Aspects of auditory processing related to speech perception. In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of*

- Phonetic Sciences*, Chapter 17, pp. 539–565. Oxford and Cambridge, Massachusetts: Blackwell.
- Moore, B. C. J. (1997b). *An Introduction to the Psychology of Hearing* (4th ed.). San Diego et al.: Academic Press.
- Norman, J. (1988). *Chinese*. Cambridge Language Surveys. Cambridge: Cambridge University Press.
- Pätzold, M. (1997). *KielDat* – data bank utilities for the *Kiel Corpus*. 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. 117–126.
- Ramsay, J. O., K. G. Munhall, V. L. Gracco, and D. J. Ostry (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America* 99(6), 3718–3727.
- Rodgers, J. E. J. (2000). Reduction rules in German spontaneous speech.
- Scheffers, M. T. M. and A. P. Simpson (1995). LACS: Label assisted copy synthesis. In *Proc. XIIIth ICPHS*, Volume 2, Stockholm, pp. 346–349.
- Simpson, A. P. (1992). Casual speech rules and what the phonology of connected speech rules might really be like. *Linguistics* 30, 535–548.
- Simpson, A. P. (1998). *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*. AIPUK 33. Kiel: IPDS.
- Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97–100.
- Wesener, T. (1999). The phonetics of function words in German spontaneous speech. In K. J. Kohler (Ed.), *Phrase-level Phonetics and Phonology of German*, AIPUK 34, pp. 327–377.
- Willems, L. F. (1987). Robust formant analysis for speech synthesis applications. In *Proc. of European Conference of Speech Technology*, Volume 1, Edinburgh, pp. 250–253.

A Composition of the databases

1) database containing prosodically and segmentally labelled material:

complete sessions: g07a, g08a, g09a, g14a, g19a, g21a, g25a, and g31a

isolated dialogues: g202a, g274a, g287a, g297a, and g306a

2) database containing segmentally labelled material only:

complete sessions: g10a, g11a, g12a, g36a, g37a, g38a, g41a, g42a

3) database containing the sum of the above data