

# **Aspects of ‘cliticization’ and style in German function words\***

Thomas Wesener

---

\*Most of the work reported here was carried out within the German Research Council project “Sound Patterns of German Spontaneous Speech” (DFG Ko 331/21–1) under the supervision of Klaus J. Kohler.



## 1 Introduction

This paper builds on work on function words in German unscripted speech presented in Wesener (1999). Whereas the latter mainly concentrated on the phonetics of specific items, the aim of the present study is to shed further light on sequences of items. The focus will lie on the production of personal pronouns depending on the position relative to a preceding or following verb, taking hypotheses in Kohler (1979) and (1991) as starting points. They e.g. state that the vowel of *ich* 'I' is likely to be dropped in *ich weiß* 'I know', but not in *weiß ich* (Kohler 1979). To take another example, the /r/-vocaloid in the pronoun *er* 'he' is likely to be monophthongal in *hat er* 'has he', but not in *er hat* (Kohler 1991). This study investigates these hypotheses on a broader empirical basis by using computerized corpora of both unscripted and read speech (IPDS 1995–1997; IPDS 1994).

One may regard certain function words such as pronouns as clitics of neighbouring words, e.g. of verbs. Abercrombie (1964) contrasts the light quality of the lateral in *feel it* with the dark quality in *feel ill*. He argues that since the lateral in *feel it* behaves as in *feeling*, there seems to be no word boundary between *feel* and *it*, *it* thus being enclitic.

In connected speech, however, function words frequently occur in longer sequences containing other function words so that several 'clitics' are attached to one word. In addition, there may be two alternative words to which a function word can be 'cliticized', i.e. the preceding and the following. In this case the function word may be regarded as either 'enclitic' or 'proclitic'. When investigating pronouns that precede and follow verbs, I therefore prefer the terminology 'preverbal' and 'postverbal' to 'proclitic' and 'enclitic', respectively. This does not exclude the possibility that certain non-standard productions may be lexicalized (see discussion).

Because of the conditions of the appointment-making scenario used to elicit the unscripted material, the first person pronouns *ich* and *wir* 'we' are particularly frequent (Wesener 1999). Other frequently occurring pronouns include *es* 'it' and *Sie* 'you'. These pronouns are thus suitable objects for investigating repetitions of the same pronoun-verb/verb-pronoun sequences. The third person pronouns *er* 'he' and *sie* 'she', on the other hand, are rare in the Kiel corpus published so far; data from a different scenario are needed to study the effects of pre- vs postverbal position on the production of these items.

A further section deals with phenomena in read speech that have already been described for unscripted speech in Wesener (1999): correlates of /x/ and /h/ and vowel nasalization instead of nasal contours. These investigations allow for further statements on the production of function words depending on speaking style.

## 2 Method

The data mainly derive from the 16 complete 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<sup>1</sup>, another the data with segmental labels only<sup>2</sup>, and a third the complete unscripted material<sup>3</sup>. These databases represent the labelling version of January 2000 and are identical to those used in the contributions by Kohler and by Kohler and Rodgers in this volume.

For the investigation of aspects of cliticization in the pronoun *er*, the above databases do not contain a sufficient number of productions due to limitations of the appointment-making scenario. Therefore, further data were elicited with a video-task scenario (Peters 2000). These data contain a significant number of relevant *er* tokens, which were labelled segmentally and entered into a *Kieldat* database.

Scripts are used to retrieve the sequences of interest, to count incidences of phenomena captured by the labelling, and to obtain durational information. These automatically generated data are complemented by an auditory/visual investigation and a narrow transcription of productions where correlates of certain phonological units are only present as modifications of neighbouring sounds and cannot be captured by a linear segment. A typical example is the temporal flexibility of glottal correlates of /h/ in German: *da haben* 'there have' can thus be realized as [d̥həm] where the correlates of the vowel of *da* and the /h/ occur simultaneously. These phenomena are often marked with the label MA in the unscripted corpus, and deleted labels can be used as points of departure for finding these cases

---

<sup>1</sup>Complete sessions g07a, g08a, g09a, g14a, g19a, g21a, g25a, and g31a; isolated dialogues g202a, g274a, g287a, g297a, and g306a.

<sup>2</sup>Complete sessions g10a, g11a, g12a, g36a, g37a, g38a, g41a, and g42a.

<sup>3</sup>Sum of the above data.

in the read corpus. Investigating them results in a more complete picture of how certain phonological units are coded phonetically, and therefore allows for a more adequate description and interpretation of certain phenomena.

In the case of pre- and postverbal pronouns, lists containing all verbs that can follow or precede the pronouns were built and integrated into scripts that extracted relevant lines from a lexicon with two-word entries.

String information from *Kieldat* databases was also used to conduct acoustic measurements in the 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, pre-emphasis factor 0.95, LPC order 16). The programs seem to work more reliably for F1 and F2 than the ESPS tool *formant* 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 improbable 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.

Frequency values were converted to perceptual units since Hertz values do not indicate whether changes are relevant in everyday communication and in the judgment of quality changes by phoneticians. A linear plotting of the acoustic formant frequencies in Hertz values would mean too high a resolution for high frequencies and too low a resolution for low frequencies. A logarithmic plotting of the frequency data, on the other hand, results in a higher resolution for low frequencies than is perceptually relevant.

The scale of ERB rate according to Moore (1997) was adopted to reflect the perceptual significance of value changes:

$$E = 21.4 \log_{10}(4.37F + 1).$$

ERBs are the equivalent rectangular bandwidths of auditory filters estimated by using the so-called notched-noise method (Patterson 1976). The scale of ERB rate has a higher resolution than the Bark scale (Zwicker 1961) for frequencies below 500 Hz, which is of relevance mainly to F0 and F1. According to Traunmüller (1990), the Bark scale is a measure of

tonotopic position, whereas ERB rate is a measure of frequency resolution and therefore seems to be more appropriate for the present study.

### 3 Personal pronouns in pre- and postverbal position

This section deals with the production of pronouns depending on their position relative to an adjacent verb. A number of frequent pronouns are selected, which allows for the investigation of repetitions of the same pre- and postverbal sequences and for quantitative statements.

#### 3.1 /r/-vocoids in *wir*

Our expectation is that the deviance from the canonical form is greater in postverbal than in preverbal productions of *wir*. Before we can address this topic, we need an overview of the acoustic properties of the /r/-vowel in *wir*. In order to know how the vowel 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* ‘how’. The results are presented in table 1. Since context is not controlled in unscripted speech, standard deviations are much higher than in experimental setups where target items are embedded into a stable carrier sentence. The high standard deviations reflect considerable variability of vowel quality.

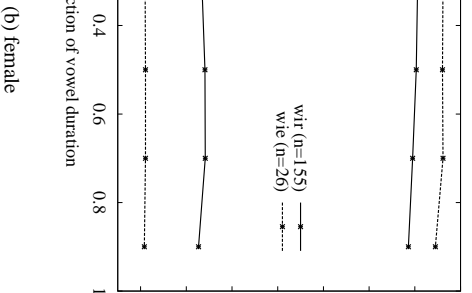
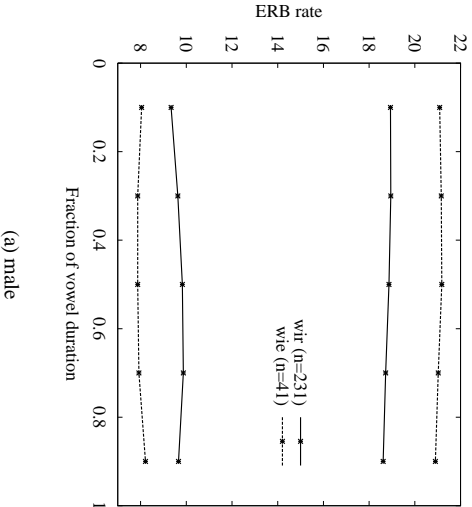
Figure 1 visualizes the data from table 1 using the scale of ERB rate (cf. section 2). Contrary to what one would expect for an opening diphthong, there are hardly any formant movements in *wir*, but the formants run almost in parallel with those for monophthongal *wie*. There are two main differences between the patterns for *wir* and *wie*: clearly lower F2 values and clearly higher F1 values for *wir*.

Although F1 and F2 are rather flat for *wir*, F1 shows a slight upward movement with a maximum at about 70% of vowel duration (cf. figure 1). This movement may be a residue of diphthongality, and in this case, the phonetic correlates of the r-diphthong would not be completely captured by the symbol [ɐ] which has often been used for ‘monophthongal’ productions. It may be tested in a perception experiment whether a slight upward move in F1 increases the *wir* percept for formant patterns intermediate between those for *wir* and *wie*. Most probably, however, the bulge is the

Table 1: Mean F1 and F2 values (in Hz) and segment durations ( $t_{obs}$ , in ms) plus 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).



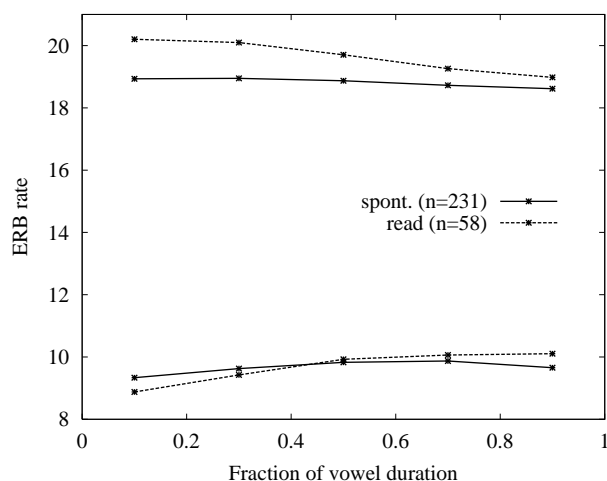


Figure 2: F1 and F2 for *wir* in read and unscripted speech (male speakers, unaccented renditions without hesitational lengthening).

product of averaging out trajectories of different shapes: while the majority of the productions displays a flat F1, some more diphthongal versions provide the slight upward tendency in the averaged curve.

In read *wir*, the formant trajectories more discernibly converge than in unscripted speech and thus result in a more diphthongal pattern (cf. figure 2). This diphthongality partly derives from a greater mean duration of the vowel in read speech. The longer the vocoid, the greater the chances are for a diphthongal production, a phenomenon already observed for utterance-final productions of *vor* by Simpson (1998).

In order to shed more light on the relevance of duration for the production of the vocoid, figure 3 contrasts productions in unscripted and read speech in the same durational classes. 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), and the three classes containing a sufficient number of productions in both unscripted and read speech are displayed. The five measurement points are projected onto the mean duration within the class (covering 80% of this duration, i.e. from

10% to 90%). This results in a pseudo-absolute representation of the durational classes that indicates acoustic change more accurately than a time-normalized representation, in which inclinations of the curves do not represent speed of acoustic change. For the sake of better comparability, the graphs for the different durational classes are aligned around their mid-points.

In spontaneous speech, 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.

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 unscripted speech, the formants are flatter and converge to a smaller degree, or even slightly diverge in the case of the shortest productions. Whereas there is a clear tendency in read speech to ‘conserve’ the transition from a close to an open vowel, the /r/-vocoid in unscripted speech manifests itself in a less linear way, with the starting and end points tending to get closer to each other.

Although there are differences in class size, number of speakers, and context distributions, the mean duration of the vowels in the 31–60 ms class in read speech is identical to the one in unscripted speech (45 ms). Different formant patterns across styles do not seem to be due to duration.

Controlling the durational parameter in the acoustic measurements reveals that non-prototypical productions (in this case formant trajectories) are not merely a consequence of temporal compression. The alignment of the speech movements is thus not completely left to the mechanics of the vocal tract, which is forced to compression by durational constraints; rather, the speaker selects a certain mode according to style.

After the investigation of the /r/-vocoid in *wir* in general, we now turn to the subject of pre- vs postverbal position and its relevance for the production of the vocoid. 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 lacks reliable information on sentence-accent, i.e. some of the tokens without a sentence-accent marker may be accented. But the patterns found in the database without

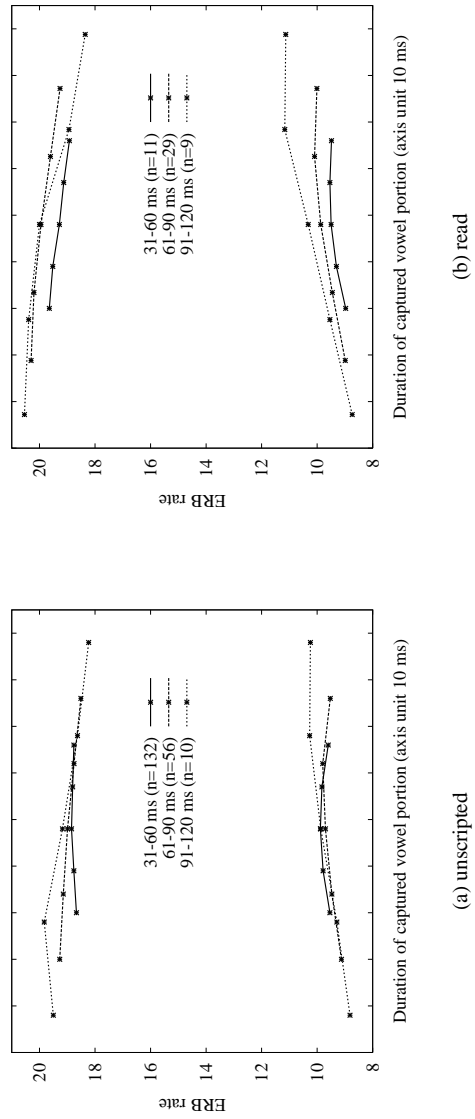


Figure 3: F1 and F2 for three durational classes of *wir* in both read and unscripted speech (unaccented renditions without hesitational lengthening).

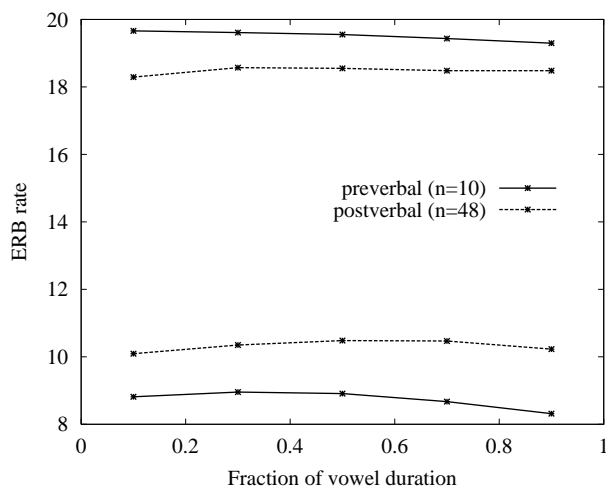


Figure 4: F1 and F2 of unstressed *wir* in *wir können* and in *können wir* (male speakers in complete database).

prosodic labels closely resemble the findings for the prosodically labelled database, and it seems legitimate to merge the databases for the present purpose.

Figure 4 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 less standard productions in postverbal position.

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

In read speech, the influence of pre- vs postverbal position cannot be investigated in sequences containing *können* and *wir* since *wir* occurs only in postverbal position (*Können wir nicht Tante Erna besuchen?* ‘Can’t we

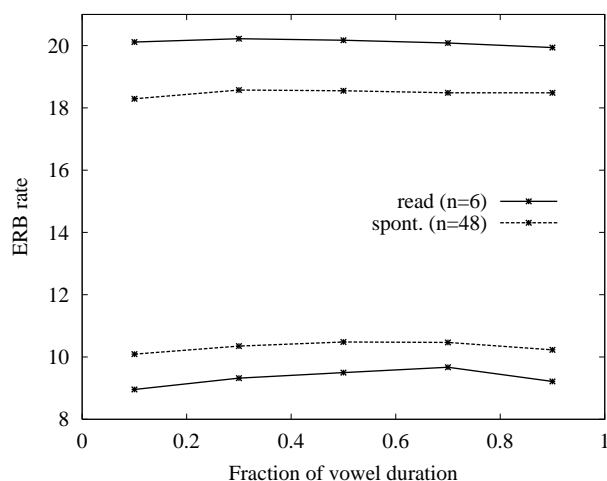


Figure 5: F1 and F2 of unstressed *wir* in *können wir* in read and unscripted speech (male speakers).

pay a visit to aunt Erna?', be054). Compared with unscripted speech, the /r/-vowel in *können wir* has a more peripheral and a slightly more diphthongal quality for F1, cf. figure 5.

### 3.2 /r/-vocoids in *er*

As mentioned in section 2, data for the /r/-vocoid in *er* derive from the video-task scenario because the appointment-making scenario does not provide for a sufficient number of productions of the item. The video-task sessions processed so far involve only two male speakers, and therefore it was decided to investigate female speakers only (sessions l01a–l04a).

Figure 6 displays F1 and F2 for these sequences in the video-task database (tokens with hesitational lengthening and sentence accent were excluded). As in *wir* above, the /r/-vocoids are rather monophthongal in both positions. The difference lies mainly in a higher F1 for preverbal productions, which points to a more open quality preverbally. At first sight, this may be seen to contradict the findings for *wir* above, where preverbal pro-

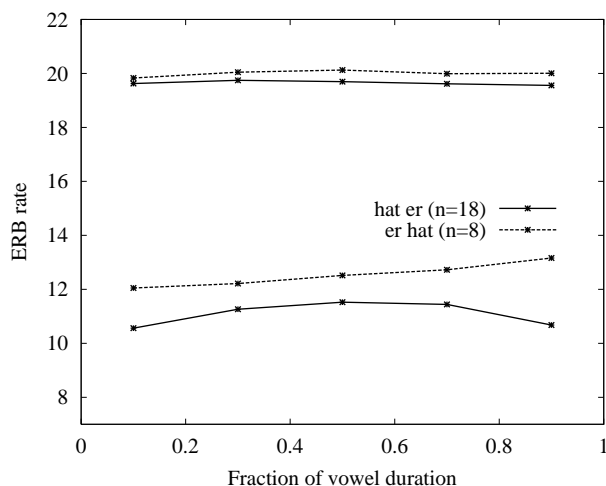


Figure 6: F1 and F2 of unstressed *er* in *hat er* vs *er hat* for female speakers.

ductions are less open than postverbal ones. But the point of departure is a comparatively close vocoid for *wir* and an open one for *er*. Therefore the less open quality in postverbal vs preverbal *er* can be understood as a sign for an articulation that more clearly deviates from a citation form pronunciation, *er* thus being in accordance with the pattern found in *wir*. Another reason for the more open quality of preverbal *er* may be that preverbal *er* lies within the same vocoid stretch as the open vowel of *hat*, /h/ being realized as a vocoid with non-modal glottal activity (breathy voice).

### 3.3 Centralization in *Sie*

Adopting a hypothesis from Kohler (1979), the monophthong in *Sie* is expected to be more central in postverbal than in preverbal position. 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 investigated.

Again, the item was more frequent postverbally, where it also occurred

with a greater variety of verb types. Figure 7 displays formant movements in pre- and postverbal position for male and female speakers in unscripted 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 8 shows the results for read speech. In accordance with our expectation, the postverbal formants lie within the preverbal formants for both male and female speakers.

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

### 3.4 Vowels and glottal activity in *ich*

Following Kohler (1979), our hypothesis is that the vocoid is more often deleted when preceding the verb than when following it. The vocoid and contoid in *ich* can be produced with the same tongue configuration: devoicing leads to a stronger airstream, which then causes frication at the hard palate. 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 voicelessness throughout *ich*, which results in a contoid only. A vocoidless production is thus more probable in phrase-initial sequences such as *ich bin* 'I am' than in *bin ich* which mostly occurs phrase-internally.

When regarding all tokens of *ich* before or after verbs in the first person singular (more than 1200 in the complete database), it turns out that the vowel in preverbal *ich* is more frequently 'absent' (7%) than postverbally (3%). In a number of these cases, however, non-segmental correlates of the vowel are present and are indicated by the label MA.

If a vowel has been segmented, a glottal reflex is found almost three

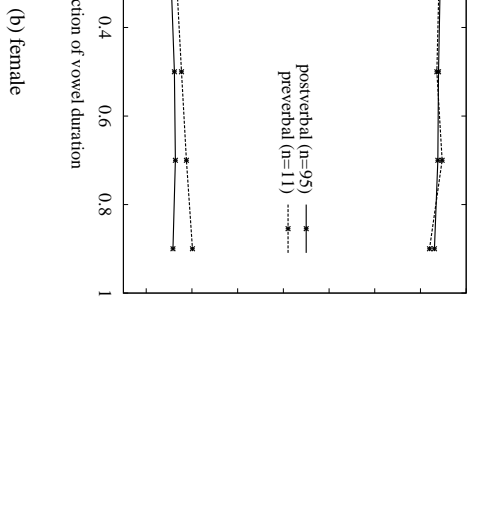
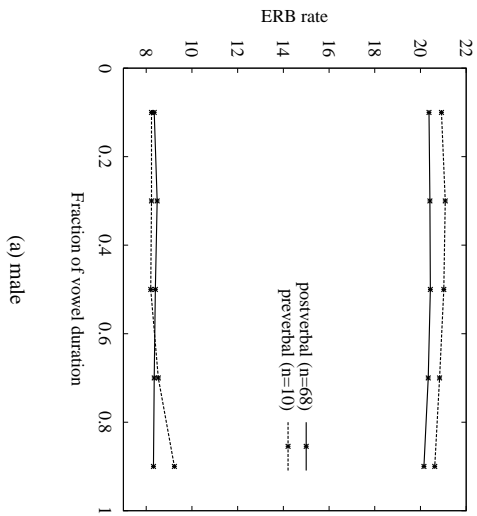


Figure 7: F1 and F2 for unaccented *Sie* in pre- and postverbal position, complete unscripted database.

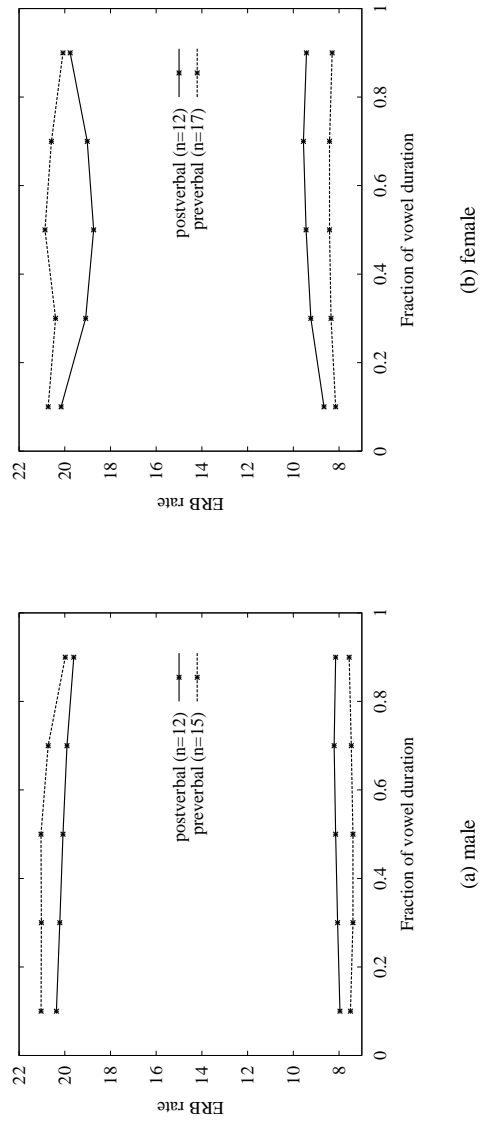


Figure 8: F1 and F2 for unaccented *Sie* in pre- and postverbal position, complete read database.

times as often in preverbal than in postverbal position (64% vs 23%). The reason is that the vowel of *ich* occurs phrase-medially in sequences such as *bin ich*, but phrase-initially and mostly after non-verbal material like pauses or breath in *ich bin*, which favours the presence of a glottal stop.

With regard to the consonant, there is hardly any difference between the positions, 'deletions' (2%) and 'replacements' by other fricatives (e.g. labiodental or alveolar ones, 1%) occur with the same frequency. Non-modal glottal activity without supralaryngeal frication, however, only occurs postverbally.

In the read corpus, which contains more than 600 sequences of pre- or postverbal *ich*, the vowel in preverbal *ich* is more frequently 'deleted' than in postverbal position (14% vs 4%), a finding that also holds for unscripted speech (see above), and that corroborates the earlier observation by Kohler (1979) on a quantitative basis.

Glottal reflexes in read vowels are also considerably more frequent in preverbal than in postverbal productions (68% vs 15%). Consonant 'replacements' and 'deletions' occur each in about 1% of the cases, irrespective of position.

Figure 9 shows an unscripted production involving a vocoid with breathy voice and no dorsal frication, which seems to occur mainly before nasals (another example being *kann ich noch* in g315a007). The glottal activity considerably overlaps the following nasal in this case. For replacements of back dorsal gestures by laryngeal adjustments before nasals in the case of velar or uvular fricatives cf. Wesener (1999), and section 4.1 below.

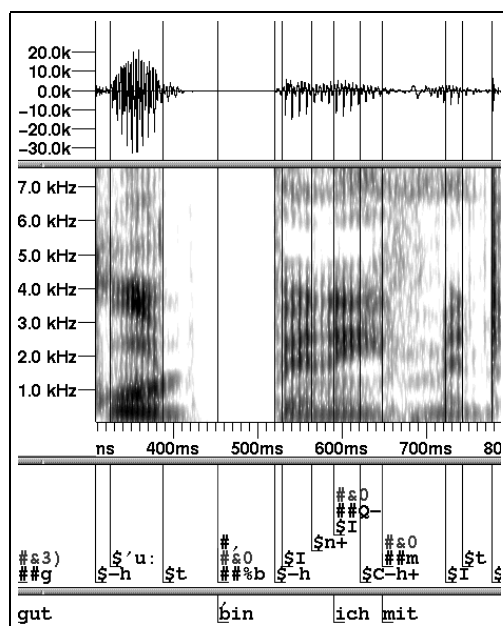


Figure 9: Breathy voice instead of dorsal activity in *bin ich mit* (g072a019).

### 3.5 Vowels and glottal activity in *es*

The third person pronoun *es* is frequently produced with an alveolar fricative only, and these forms are often indicated with 's in German orthography, including the transliteration of the Kiel corpora. Although single productions of 's are ambiguous in that they could also derive from *das* 'that', it seems legitimate to combine the two orthographic forms *es* and 's and think of them as both belonging to the lexical item *es*. In order to gain insight into the relevance of pre- vs postverbal position for the production of *es* and 's, the orthographic items were investigated before and after third person singular verbs in the complete unscripted database.

As with most other pronouns, *es* occurs more frequently after (277) than before verbs (61). Vowels are often absent in connection with a subset of frequent verbs that are mostly function words themselves, e.g. *ginge*

'would suit', *ist* 'is', *wäre* 'would be', *war* 'was', and *wird* 'will'. In postverbal position, there is no vowel in 40% of the cases, whereas before verbs vowels are absent in less than a quarter of the tokens.

A clear difference between post- and preverbal positions is also found in the production containing a vowel. The absence of a glottal reflex at the vowel onset, labelled as Q-, occurs in three quarters of the postverbal productions and is almost twice as frequent than in preverbal position.

In read speech, realizations without vowels also occur more frequently after than before verbs. Again, lack of glottal activity is much more frequent in postverbal vs preverbal *es*. The difference between the positions as to glottal activity is even greater in the read than in the unscripted corpus since all preverbal tokens of *es* in read speech occur utterance-initially, thus favouring glottal activity.

## 4 Other phenomena in read speech

In this section, phenomena that have already been presented for unscripted speech in Wesener (1999) are investigated for read speech. It is discussed whether a broad range of non-standard productions is also found in read speech, and if these realizations occur with the same frequency.

### 4.1 /x/

In unscripted speech, nearly a fifth of the tokens of *noch* 'still', *auch* 'also', *doch* 'yet', and *nach* 'to' are marked with either x- or x-h. The percentage for read speech is considerably lower (6% of more than 1000 occurrences). Although this might in part be due to the more frequent use of changed labels in the segmentation of the unscripted corpus, these numbers indicate that non-canonical productions are less numerous in read speech.

The range of non-canonical phenomena connected with x is nonetheless wide. Dorsal articulation can be given up in favour of non-modal glottal activity, particularly breathy voice: *noch sagen* 'still say' [nɔ̃fizaʊp̥] (d1me054), *auch noch* [aʊfɪnɔ̃x] (hpt̥e052), similarly *nach Mannheim* (kkoe062). Breathy voice for /x/ can also be shifted; in the case of *noch die Verbindung* 'still the connection' [nɔ̃d̥f̥-] (d1me010) the correlate of /x/ seems to be produced as late as in the vowel of the following word. Breathiness does not seem to be merely induced by the voicelessness of

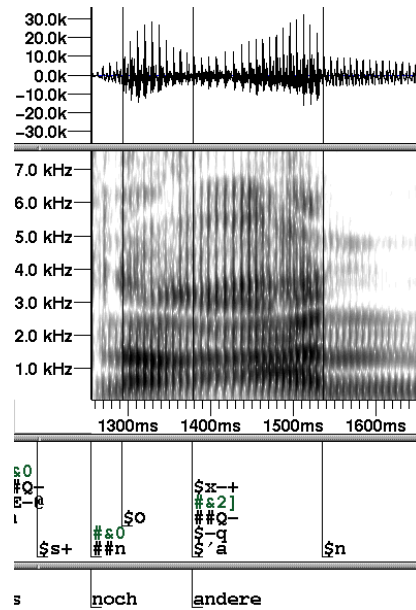


Figure 10: Correlates of /x/ in *noch an[dere]* (kkœ061).

the following labiodental fricative here because the spectral structure is more blurred at the beginning of the vowel than immediately before the fricative.

Interaction between vowel- and /x/-related glottal activity is found in *noch andere* 'still other' (cf. figure 10, kkœ061) where /x/ seems to be marked by less spectral energy as well as a slight drop in F0. These cues may also partially be the correlates of the glottal activity connected with the following vowel. In *auch eine* 'also a' (cf. figure 11, ugae073), the end of the diphthong in *auch* is produced with creaky voice. This stretch of creaky voice seems to be too long to simply announce the short glottal stop at the beginning of *eine*. It is probably also related to the word-initial vowel of *auch*, and its right-shift relative to the beginning of the vowel may be a correlate of /x/.



Sometimes the dorsal fricative is present, but the borderline to a following labiodental fricative is not easy to detect: *noch von* 'still from' (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.

In a complex case of *auch schon* 'also already', the postalveolar fricative has a formant around 1200 Hz, and its energy cut-off lies as low as 700 Hz (cf. figure 12, k22but t2). Its two surrounding segments, i.e. the fricative in *auch* and the vowel in *schon*, have been marked deleted. Both the fricative and the nasal in *schon* are velarized and thus sound darker than the fricative-nasal sequence in e.g. *rauschen* 'rustle'. This velarization partly seems to be a non-sequential correlate of the 'deleted' vowel (marked by MA), but it is probably stronger than would be necessary for the vowel alone; there might even be a primary postalveolar-velar double articulation in the fricative. In addition, the fricative is preceded by a breathy-voiced stretch which together with the velar component of the fricative seems to represent /x/.

One case where x has been deleted in *doch* displays short dorsal friction in the transition to the velar stop in *doch gleich* 'yet soon' (k69mr-092). The dorsality of the following plosive seems to allow for a shortened production of the dorsal fricative here.

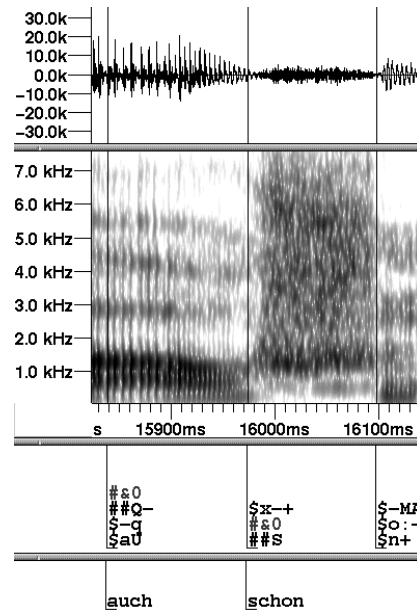


Figure 12: Breathy voice and the velar component of the following postalveolar fricative as correlates of /x/ in *auch schon* (k22but t2).

We have seen that read speech can display quite radical 'deviations' from standard back dorsal fricatives, showing a range of phenomena that is comparable to the one found in unscripted speech. The main difference between the speaking styles is that these forms seem to occur less frequently in read than in unscripted speech.

## 4.2 /h/

h has been marked deleted in only 2% of the cases, as opposed to 9% in unscripted speech. This points towards more care with regard to the presence of breathiness for /h/ and its temporal alignment, probably induced by orthography.

The item *Bahnhof*, however, is frequently produced without breathiness 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* 'yard' is opaque. The frequent lack of non-modal glottal activity seems to indicate that *Bahnhof* is interpreted as mono- rather than bimorphemic. This tendency is probably even stronger in unscripted speech, where the speaker is not influenced by orthography; unfortunately, the item does not occur in the present corpus of unscripted speech.

How is word-initial /h/ realized after vowels in read speech? Unfortunately, the sequence *da haben* 'there have' that frequently occurs in unscripted speech (cf. Wesener 1999) 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 unscripted speech.

## 4.3 Vowel nasalization in *uns* vs *und*

In read *uns* 'us', the nasal contour is always present ( $n = 64$ ). The amount of productions without contour for *und* 'and', however, is very similar to

the one found in unscripted speech (1%,  $n = 551$ ). For read *uns*, the orthographic presence of <n> seems to block the tendency for realizations without nasal contours, which in unscripted speech is much stronger for *uns* than for *und* (cf. Wesener 1999). The result may be interpreted as suggesting that orthography has a conserving effect on pronunciation.

## 5 Discussion

The results on the influence of pre- vs postverbal position have to a large extent corroborated the qualitative observations in (Kohler 1979), using quantitative data from large corpora of unscripted and read speech. To take the pronoun *wir* as an example, acoustic measurements in the corpora have shown that it indeed is produced with a more open vowel quality after than before verbs.

Combinations of verbs and postvocalic pronouns such as *wollen wir* 'shall we' can become close to lexicalized as single units. The name of the German comedy group *Samma Womma Nomma* stands for *Sag' mal, wollen wir noch mal?* 'Tell me, shall we once again?' Besides reflecting the lack of velar and lateral activity in *sag' mal* and *noch mal*, it also demonstrates the quasi-enclitic character of *wir* in *wollen wir*.

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.

Contrary to what one might expect, productions in read speech can be as remote from isolated word forms as in unscripted speech. The main difference between the speaking styles seems to be that these productions occur more frequently in unscripted than in read speech (cf. e.g. the results for word-final /x/).

The findings, besides addressing the topics of pre- vs postverbal position and style, also have consequences for phonological interpretations of the phonetic data. From a procedural point of view, one might postulate that the vowel preceding /r/ is 'elided' in monophthongal productions of

wir: [ɪɐ] → [ɐ]. Both phonological units, however, are phonetically coded throughout the monophthong: F1 is intermediate between the endpoints of the diphthongal productions (cf. figure 3). A non-derivational understanding of phonetic properties as direct exponents of phonological units seems more appropriate to account for these data.

In a cross-language perspective, the variation of dorsal fricatives with breathy voice and breath in German is found as an alternation in Czech voicing assimilation (Dankovičová 1999). In 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, and probably merely glottal articulations can be found in connected speech. Temporal flexibility of glottal correlates for /x/ and /h/ is probably not exclusive to German, but may also be found in the other languages mentioned above.

From a diachronic point of view, Czech has lexicalized the production of [ɸ] for velar articulations that occur in the cognates of other Slavic languages (Olle Engstrand, personal communication). In a similar vein, Germanic /h/ in German *Horn* and English *horn* corresponds to /k/ in Latin *cornus*. The Germanic forms seem to derive from intermediate productions with dorsal fricatives (synchronically, such fricative productions of dorsal plosives are found e.g. in Tuscan varieties of Italian).

Findings from corpora of connected speech can thus serve to form hypotheses about connected speech phenomena in other languages, and also about sound change. Although corpora of read speech display a number of interesting sentence-level phenomena when closely investigated, the analysis of unscripted speech with its more frequent 'deviation' from isolated word pronunciations is an even better starting point for understanding human speech at the level of sentences and utterances.

## 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.
- 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.
- IPDS (1994). *The Kiel Corpus of Read Speech*, Volume 1, CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- IPDS (1995–1997). *The Kiel Corpus of Spontaneous Speech*, Volume 1–3, CD-ROM#2–4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- 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. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* 48, 180–192.
- Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Moore, B. C. J. (1997). 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.
- Norman, J. (1988). *Chinese*. Cambridge Language Surveys. Cambridge: Cambridge University Press.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America* 59, 640–654.

- 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.
- Peters, B. (2000). ‘Video Task’ oder ‘Daily Soap Szenario’: Ein neues Verfahren zur kontrollierten Elizitation von Spontansprache. URL [http://www.ipds.uni-kiel.de/pub\\_exx/bp2001\\_1/Linda21.html](http://www.ipds.uni-kiel.de/pub_exx/bp2001_1/Linda21.html).
- 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. 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 *Proceedings of the European Conference on Speech Technology*, Volume 1, Edinburgh, pp. 250–253.
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *Journal of the Acoustical Society of America* 33, 248.