

Accounting for the phonetics of German *r* without processes

Adrian P. Simpson

*Institut für Phonetik und digitale Sprachverarbeitung
der Christian-Albrechts-Universität zu Kiel*

1 Introduction

A rich variety of phonetic patterns are associated with German *r*. Leaving aside interdialectal and interstylistic differences, these patterns are determined by the place in the syllabic and rhythmic structure. For a typical North German speaker the correlates of *r* can range from a voiceless uvular fricative in the initial consonant cluster of a word such as *trat* ("stepped") to apparent absence of anything at all following the open vowel in a word such as *Bart* ("beard"). In particular, the patterns observed for *r* in postvocalic position are particularly rich. In combination with short quantity vowels (e.g., *wird* "becomes, will", *Wurst* "sausage", *Korb* "basket", *Erna* proper name) we can find phonetically long monophthongs, whose quality is opener and more central than the quality of their *r*-less congeners. In combination with non-open long quantity vowels (e.g., *ihr* "her", *Uhr* "clock", *wer* "who") we can find phonetically long diphthongs which end centrally somewhere between [ə] and [ɐ].

Both in extensive descriptive surveys (Ulbrich 1972; Graf and Meißner 1996) as well as phonological analyses (Hall 1993), these patterns are accounted for in terms of generative processes. The names of the descriptive categories 'vokalisiert' and 'elidiert' used in Ulbrich (1972) and Graf and Meißner (1996) are process-oriented. In Hall's lexical phonological account all allophonic variants are derived from the consonantal specification of a voiced uvular trill employing rules such as '[R]-vocalisation' (Hall 1993: 88).

There would appear to be both formal linguistic as well as phonetic grounds why a generative phonological account of the phonetic patterns associated with German *r* is inappropriate. In non-linear approaches to accounting for phonetic patterns such as Firthian phonology (Firth 1948) and more recently in related declarative frameworks (Coleman 1994; Local and Ogden 1997) as well as articulatory phonology (Browman and Goldstein 1989), it has been successively shown that differences in the phonetic appearance of the same phonological objects can be accounted for using a combination of rich phonological structure and non-linear phonetic exponency, avoiding the need for the destructive might of rewrite rules. From a phonetic point of view a different interpretation of the cases of elision reported in Ulbrich (1972) and Graf and Meißner (1996) suggest that the phonetic correlates of the phonological object *r* are not absent.

Using a declarative phonological analysis and non-linear phonetic exponency this paper demonstrates that the complex set of consonantal and vocalic patterns associated with *r* can be reduced to two phonetic exponency statements, without using processes such as vocalization and elision. One exponency statement describes the consonantal correlates associated with *r* at syllable onset. The second exponency statement describes the vocalic correlates of *r* at coda. The monophthongal vocalic qualities found in connection with short quantity vowels as opposed to diphthongal patterns found with long quantity vowels are seen as the product of differences in co-temporality of the correlates of *r* and those of the vowel. However, these differences are not treated as being specific to *r*, but rather as part of the more general observation that consonantal strictures following short vowels are often longer than those following long vowels. Verification is provided using synthesis examples produced by a computer implementation of the phonological abstractions and their phonetic exponents.

2 The phonetics of German *r*

Comparing descriptions from around the end of the last century (Bremer 1893; Viëtor 1894) with contemporary analyses (Kohler 1995), it is striking how little has changed in the complex set of phonetic patterns associated with *r* in Standard German.

The consonantal correlate of *r* is a dorso-uvular stricture of close or open approximation. The state of the glottis accompanying the stricture can be open, narrowed or vibrating. The state of the glottis and the type of stricture are partly contextual and partly speaker-specific. In voiceless plosive and fricative onsets the glottis is open and a dorso-uvular stricture of close approximation gives rise to friction. In other onsets and intervocalically the glottis is ready for voice¹, but if the dorso-uvular stricture is too small, the build-up of air pressure between the glottis and the supraglottal stricture can be sufficient to suppress vocal fold vibration (Bickley and Stevens 1986, 1987; Stevens 1987).

Figure 1 contains sonagrams and annotations of utterance portions illustrating these different consonantal possibilities². Example 1(a) is from a male speaker, the remaining examples are from female speakers. The portion labelled with \$*r* in each case is of interest. The first three examples (1a-c) show voiceless uvular friction from voiceless plosive (1a) and fricative (1b-c) onsets, taken from the words (a) *Eintracht* ("harmony"), (b) *schreiben* ("write") and (c) *Freitag* ("Friday"). The uvular friction in each case is characterized by strong excitation of F2-F4, most clearly visible in the female examples (1b-c). Figure 1(e-f) illustrates unvoiced (e) and voiced (f) uvular friction in two tokens of the proper name *Doris* uttered by the same speaker. As both tokens are temporally very similar and were produced by the same speaker, the presence or absence of vocal fold vibration would appear to arise solely from differences in the size of dorso-uvular stricture. The most complex glottal activity during the uvular stricture arises in lenis plosive onsets. The utterance portion annotated with \$-h and \$*r* in 1(d) is a typical example. Following the release of the velar plosive the fricative stricture is unvoiced, then after about 30 ms voicing begins, only to be almost completely suppressed again after a further 20 ms. Similar complexity can also be found following labial and apical plosives. This complexity arises from the instability of uvular strictures, which are particularly susceptible to abrupt changes in air pressure and flow occurring directly after plosive release.

At coda the phonetic correlate of *r* is a central half-open vowel quality. However, in the majority of cases this quality is not temporally delimitable in the acoustic record in the same way as the dorso-uvular fricatives and approximants just described. Instead we find a range of monophthongal and diphthongal vowel qualities, which are temporal amalgams of the phonetic correlates of *r* and those of the vocalic nucleus. Monophthongal [ɐ] for the weak syllable *ər* is well-known from the literature (Meinhold 1989; Kohler 1990; Kohler 1995; Barry 1995), but in combination with other vowels we are led to expect diphthongal vowel qualities which begin at the quality of the *r*-less vowel and move towards a [ɐ]-quality. The descriptive dichotomy of a monophthongal [ɐ] for *ər* and diphthongs in combination with other vowels is largely based on observation of isolated words and syllables and oversimplifies the patterns found even in

¹Cf. Lisker and Abramson 1964, p. 415: 'If the speaker closes the glottis down enough for phonation, he does not directly "command" the vocal folds to vibrate; rather, he makes the necessary muscular adjustments that set the conditions for vibration when sufficient airflow is supplied.'

²All the examples in this paper are taken from speakers who produced the Marburg and Berlin sentence set from the *Kiel Corpus of Read Speech* (Kohler, Pätzold, and Simpson 1995; IPDS 1994). The index of each example (e.g. k08mr074) refers uniquely to an utterance by a particular speaker from a particular subcorpus. The index 'k08mr074' in Figure 1, for instance, refers to sentence 074 of the Marburg sentence set spoken by speaker k08 (uneven numbers are male speakers, even female). The natural and synthetic utterances contained in the figures and tables in this paper can be found at the following URL: www.ipds.uni-kiel.de/examples.html.

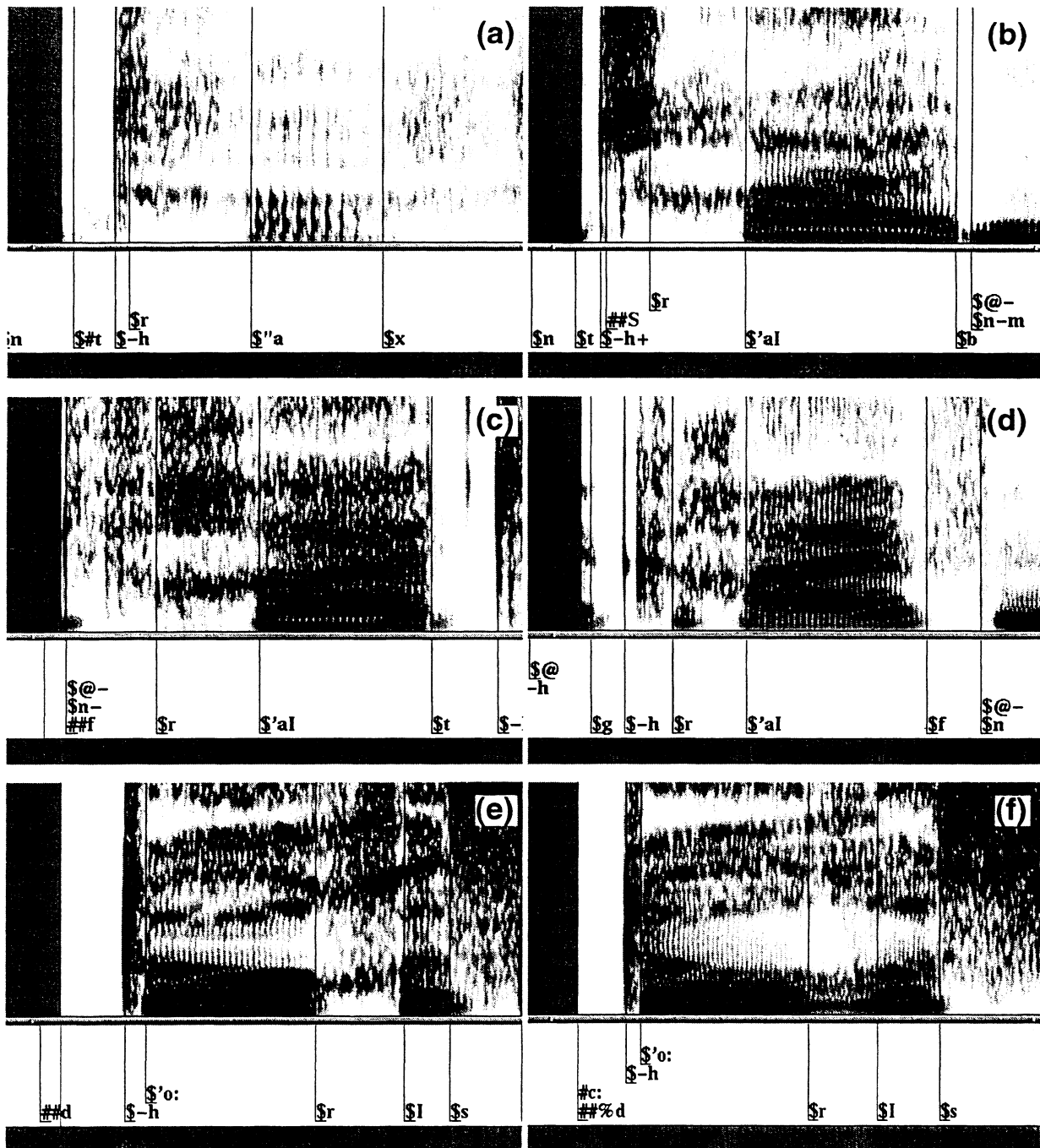


Figure 1: Sonograms and annotations illustrating different strictures and states of the glottis associated with the dorso-uvular correlate of *r*. Examples (a-c) are voiceless uvular fricatives found in voiceless plosive and fricative onsets, (d-f) are from other onset and intervocalic positions (see text). Examples (b-f) are from female speakers, (a) from a male speaker. (Refs.: (a) k07mr055, (b) k10mr095, (c) k10mr063, (d) k12mr027, (e) k08mr026, (f) k08mr074)

laboratory read speech. In long quantity syllables the phonetics of the vowel and *r* give rise to diphthongs beginning at a quality akin to the *r*-less vowel and ending centrally between [ɔ] and [ɐ]. In short quantity syllables the phonetics of the vowel and *r* produce monophthongal or slightly diphthongal vocalic portions, whose quality is open and central of the corresponding *r*-less vowels.

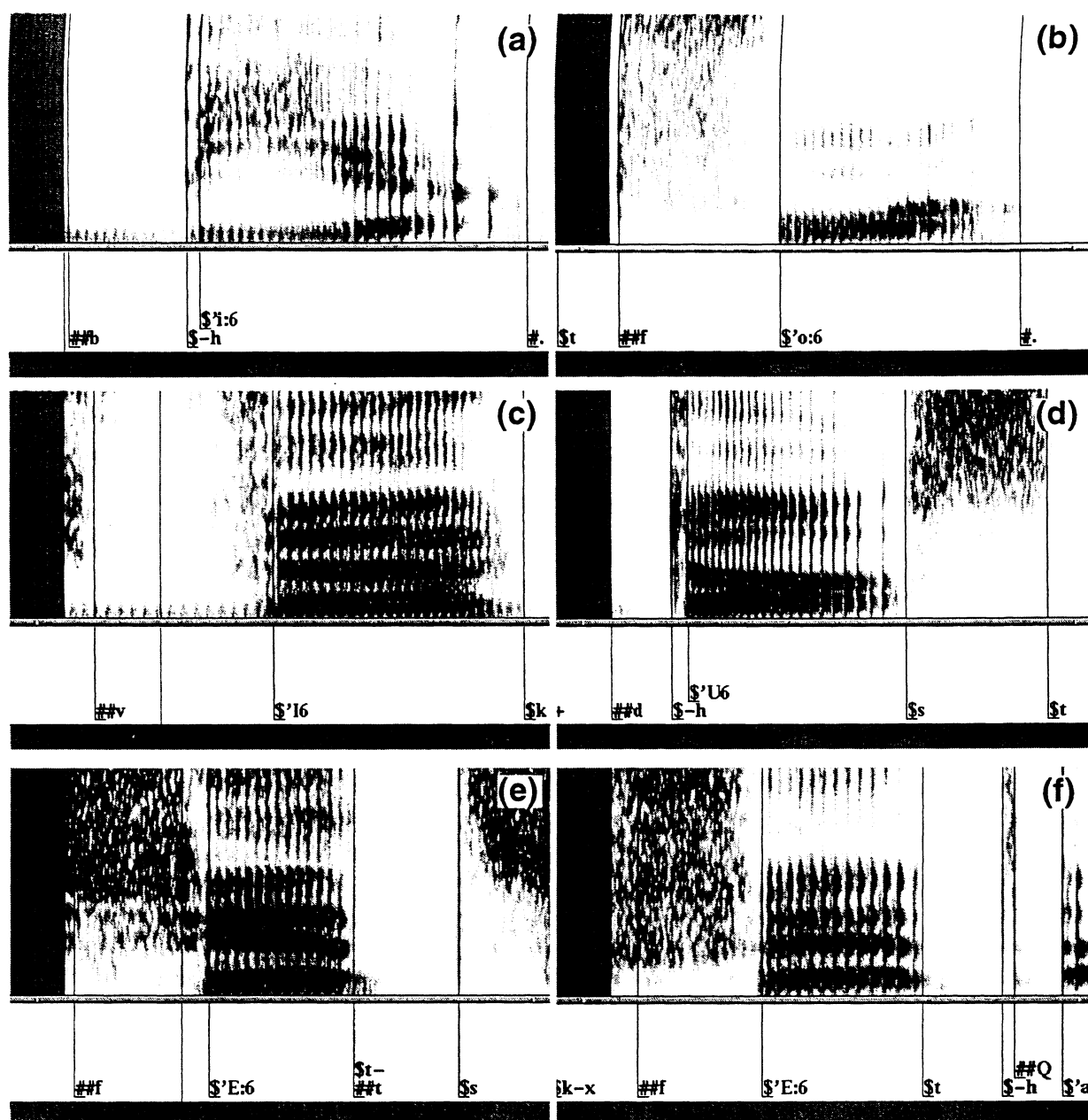


Figure 2: Sonagrams and annotations of a selection of short and long quantity r-vowels produced by the male speaker k67. Utterance portions are from the words (a) *Bier* ("beer"), (b) *vor* ("before"), (c) *wirklich* ("really"), (d) *Durst* ("thirst"), and (e-f) *fährt* ("goes, drives"). (Refs.: (a) k67mr089, (b) k67mr058, (c) k67mr090, (d) k67mr062, (e) k67mr026, (f) k67mr071)

Figure 2 contains sonagrams and annotations of short and long quantity r-vowels. Figure 2(a) and (b) are examples of long quantity vowels, 2(c) and (d) short quantity. The long quantity vowels in 2(a) and (b) are both utterance final and the diphthongal formant movements from the vowel space periphery to a central half-open position in both cases is clearly visible. Indeed, the beginning of the syllabic portion in 2(a) is voiced dorso-palatal friction. In stark contrast to the clear diphthongs in 2(a) and (b) are the short quantity vowels in 2(c) and (d). The vocalic portions here are also relatively long, ca. 150 ms for *wirklich* ("really") and 200 ms for *Durst* ("thirst"). However, in both cases the auditory quality of the vocalic portions is monophthongal, which is reflected in the presence of any significant formant movements only in the transitional

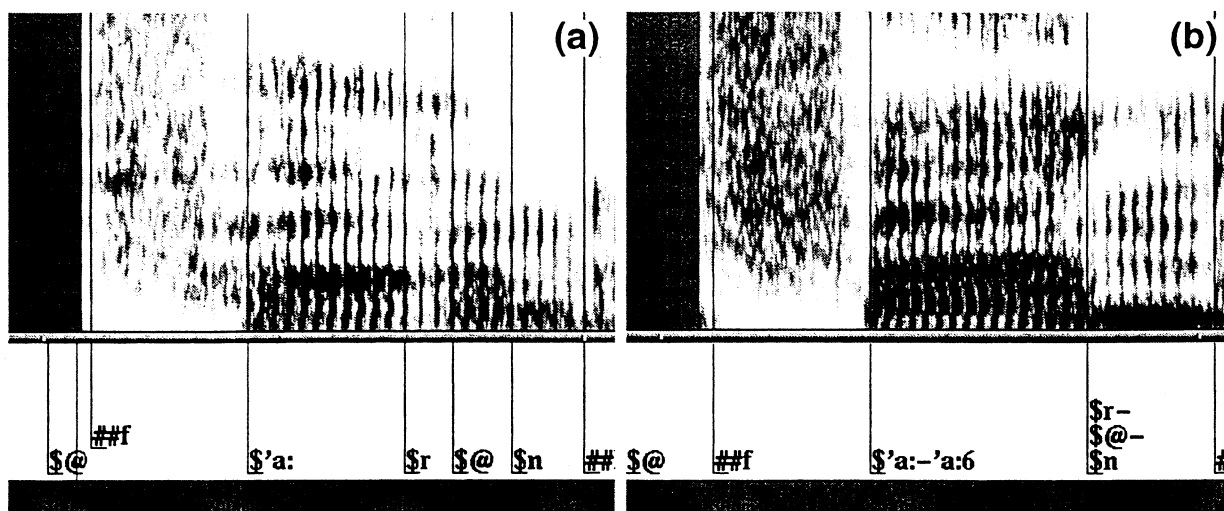


Figure 3: Sonagrams and annotations of (a) consonantal and (b) vocalic tokens of the verb *fahren* from two male speakers. (Ref.: (a) k07mr088, (b) k11mr088)

periods away from and into adjacent consonants.

Two further aspects complicate the description of *r*-vowels. First, different tokens of the same word by the same speaker can have different vowel qualities. Tokens of the word *fährt* (“goes, drives”) in Figure 2(e) and (f) illustrate one such example. The word *fährt* occurs twice in the Marburg sentence set, in *Doris fährt zu weit links* (“Doris is driving too far to the left.”) and *Vorsicht, Zug fährt ab!* (“Mind out, the train is departing!”). Consistently across the twelve speakers who produced this sentence set, the vocalic portion of the first token is qualitatively closer and more likely to be diphthongal than that of the second token. It is not clear on the basis of the data base material used whether this is categorial ambivalence or due to long domain vowel harmony. The systematic nature of the difference points towards harmony. In the first case the vowel of *fährt* is surrounded by vowels which are half-close and close in quality, whereas in the second sentence the *fährt* is followed by the open vowel of the verbal particle *ab*. The second complication are differences in the quantitative distribution of *r*-vowels across different lexical and grammatical items. So, for instance, the grammatical item *wer* (“who”) for different speakers can have either the half-open quality of the short quantity vowel or the half-close and markedly diphthongal quality of the long quantity vowel.

In the majority of cases the distribution of consonantal and vocalic correlates of *r* is clear-cut. However, the phonetic shape of certain lexical items can alternate between the vocalic and consonantal correlates. This alternation most commonly occurs in a **Vrən** configuration found primarily in certain verb forms (e.g. *fahren* “go, drive”) and plural nouns (e.g. *Erdbeeren* “strawberries”). Figure 3 shows (a) consonantal and (b) vocalic tokens of the verb *fahren* from the same sentence uttered by two male speakers (k07 and k11). In 3(a) vocalic portions are visible on either side of the voiced dorso-uvular fricative. In 3(b) an open vocalic portion extends from the labiodental friction to the onset of the final nasal. However, both tokens have approximately the same duration, and are both disyllabic. In 3(b) both the nasal and the open vocalic portion are longer and the nasal is syllabic. In the next section structural differences will be proposed to account for the consonantal and vocalic tokens, but it is unclear whether a speaker chooses to produce one or the other variant purely on stylistic grounds or whether structural ambivalence may also play a part.

3 Accounting for the phonetics

Recent attempts at accounting for the complex patterns associated with German *r* have been in generative phonological terms. This applies not only to generative phonological analysis proper (Hall 1993), but also to Ulbrich's (1972) extensive descriptive survey. In the introduction both phonetic and phonological difficulties with a generative approach were identified.

The phonetic problem is one of data interpretation which is not exclusively generative, but is undoubtedly nurtured by generative formalism. It can best be illustrated using the descriptive categories which Ulbrich uses to classify vocalic allophones. Ulbrich analyses some 11000 /r/ allophones taken from recordings of news broadcasts, programme announcements and literary texts produced by 25 radio announcers and 15 actors. Each allophone is classified according to auditory and structural criteria. Five primary articulatory categories are proposed ("r-trills", "r-fricatives", "r-vowels", "r-elision" and "r-indifferent"). The first three of these categories are then further subdivided. So, for instance, "r-fricatives" is divided into [ʁ, ʁ̥], [ɹ] and [χ, x].

The most problematic aspect of Ulbrich's analysis is his categorization of vocalic allophones. Whereas the classification of consonantal allophones pays attention to articulatory and phonatory detail the categorization of the vocalic allophones is coarser and in places confusing. The "r-vowels" category is divided into [ɐ] and [V^ɐ] for monophthongal [ɐ] cases and diphthongs, respectively. Given the enclosure in [] we would expect the classification of a vocalic allophone as [V^ɐ] to mean a diphthong whose quality ends at a half-open central position, but the description in places shows this not to be the case, e.g.

... *durch* zwischen [duɐç] und [duɐç].
([ɐ] tendiert in diesem Falle sehr nach [ə] oder [ɪ])
... *durch* between [duɐç] and [duɐç].
([ɐ] in this case has a strong tendency towards [ə] or [ɪ])

(Ulbrich 1972: 93)

More confusing still is the category "r-elision" which should be reserved for those cases in which the phonetics of *r* are no longer deemed to be present. However, elision is also used to cover those cases in which the phonetics of *r* are qualitatively and/or durationally present, but cannot be temporally delimited from the phonetics of the vowel (see Figure 2c-d). Ulbrich's relatively simple classification of the vocalic allophones, then, does not reflect a simpler situation than found for the consonantal allophones, but rather an oversimplification in the description. A potentially more serious problem, however, is the absence of any criteria for establishing whether *r* is phonetically present in an utterance portion. The category "r-indifferent" is used to classify cases where a decision between elision and vocalization could not be made. But the problem is not merely the lack of operational criteria, but has theoretical implications. The lack of any qualitative or durational differences in the vocalic portions of a word pair such as *Bart* ("beard") and *bat* ("offered") is not sufficient grounds for claiming that *r* is not phonetically present in *Bart* in exactly the same way as it is in a close vowel environment such as that illustrated in Figure 2(a).

The formal problem with the generative phonological approach to accounting for phonetic patterns is that its rewrite formalism is too powerful. Coleman (1994) argues that despite repeated attempts at restricting this power, the generative formalism of transformational grammar may still represent no more than an unrestricted rewrite system. The constraint-based approach

of declarative phonology (e.g. Coleman 1994; Scobbie 1993) counters this problem by drastically reducing the mechanisms which can be used to manipulate linguistic structures to unification, which can only combine linguistic structures without changing or removing informational content. In generative phonology the rules derive the phonetics from the phonology by successively modifying and deleting structural information. In a declarative approach the phonological and phonetic levels of abstraction are kept apart and the path between the two is mediated by exponency statements which give the phonological structure a phonetic interpretation. The strict segregation of the phonetic and phonological levels of abstraction and the use of phonetic exponency to mediate between the two levels are of course central features of Firthian prosodic phonology (Firth 1948; Henderson 1949).

Although articulatory phonology (Browman and Goldstein 1989) differs in many respects from declarative phonology, not least because articulatory phonology does not have distinct phonetic and phonological levels of abstraction, it shares one important feature of interest, which is the condition that a gesture cannot be removed or added.

The stark reduction in the manipulative power of both articulatory and declarative phonology not only has formal consequences, but also has implications for the way in which we interpret and account for phonetic data. If linguistic material can no longer be deleted, but the phonetics of a particular phonological object appear to be absent we are forced to consider any one of a number of alternative accounts:

- The phonetics are there, but insufficient attention has been paid to detail.
- The phonetics are there, but are “hidden” behind the phonetics of other objects, and require other recording techniques to make their presence visible.
- The phonetics are there, but are so similar to the phonetics of another object, with which they are cotemporal, that they are not observable regardless of observational detail or recording technique.

All of these have been brought to bear in support of phenomena which generative phonology has dealt with in terms of deletion. Kelly and Local (1989) provide many examples which illustrate that detailed phonetic observation can reveal difference where identity had previously been assumed or where phonetic material which might otherwise have been considered to be absent. X-ray investigation has revealed the presence of lingual activity which is not observable in the acoustic record because it was overlaid by labial closure (Browman and Goldstein 1990). Most controversial of all is the last alternative because the presence of the phonetics is not claimed on the basis of patterns which can be observed whichever method of recording is used. However, there are a number of ways of justifying different phonetic ingredients despite surface identity. Our justification is based on a speaker for whom the vocalic portions in the word pair *Fahrt* (“trip”) and *bat* are not observably different³:

- The phonetics of *r* in other parts of the *fahr*-paradigm are observable, and can be assumed to be present in *Fahrt* as well.
- In other varieties and for other speakers of the same variety, the phonetics of *r* in *Fahrt* are observable, i.e. the vocalic portions of *Fahrt* and *bat* are different.
- A model which reproduces the observable patterns in words such as *Bier* and *Durst* accounts equally well for the vocalic portion in *Fahrt*.

³At present this example is hypothetical as the data base material does not provide suitably comparable material.

Let us summarize the discussion up to this point. We have cast doubt on certain aspects of Ulbrich's descriptive categorization, in particular the treatment of vocalic patterns subsumed under "elision". A generative account of the phonetic patterns has also been rejected on formal grounds.

A declarative, Firthian approach to accounting for the patterns described in the previous section will now be outlined. This involves proposing abstractions at the phonetic and phonological levels of abstraction and deciding which aspects of the phonetic patterns are to be accounted for at which level.

Figure 4 illustrates structural requirements at the phonological level. The different phonetic correlates of *r* are related to different places in the structure of the syllable which in general phonetic terms can be stated as follows:

- r* at onset: dorso-uvular stricture of close/open approximation
- r* at coda: half-open, central vowel quality

These exponency statements differ little from allophonic statements with the difference that these exponents define the ingredients of utterance and not what is temporally delimitable or directly observable in utterance (cf. Browman and Goldstein's 1992, 'input' and 'output' phonetics). Figure 4 shows that different affiliations of *r* to the syllabic structure can be used to account for tokens of words *fahren* which can exhibit either the consonantal (4b-c) or vocalic (4a) correlates. Indeed, if ambisyllabicity is admitted as a possible structural configuration then tokens of *fahren* with the consonantal correlate can be seen as having *r* at the coda of the first and/or at onset of the second syllable. In the ambisyllabic case (4b) both the vocalic and the consonantal correlates of *r* would be present, in the simple onset case (4c) only the consonantal correlate. In open vowel cases this difference may be difficult to verify, but with other vowel qualities (e.g. *spazieren* "stroll") ambisyllabicity would predict a more diphthongal vocalic portion in the second syllable.

The differences between monophthongal (2c-d) and diphthongal (2a-b) vocalic portions are seen in terms of differences in the amount of temporal overlap between the phonetic correlates of the vowel and those of *r*, being greater in short than in long quantity syllables. It might be the case – the model presented in the next section implements this – that the greater temporal extent of *r* in short quantity syllables is similar to that found for other consonants following short and long vowels, as has been occasionally reported for other languages (e.g. Nooteboom 1972).

As was clear from the description in the previous section certain aspects of the articulatory and phonatory behaviour associated with *r* are not to be attributed directly to the phonetic correlates of *r*. The presence or absence of vocal fold vibration in certain cases was considered to be a product of the complex interaction between articulatory configuration and air flow/pressure. The absence of voice in voiceless onsets may also in part be due to this interaction, but it is primarily voicelessness as a correlate of such onsets coincident with the dorso-uvular stricture which gives rise to voiceless friction.

In the account proposed in this section, differences in the phonetic patterns associated with *r* are attributable to:

- place in syllable structure (onset, coda);
- extent of temporal overlap of the phonetic correlates of *r* with those of other objects;
- articulatory-aerodynamic behaviour of the vocal tract.

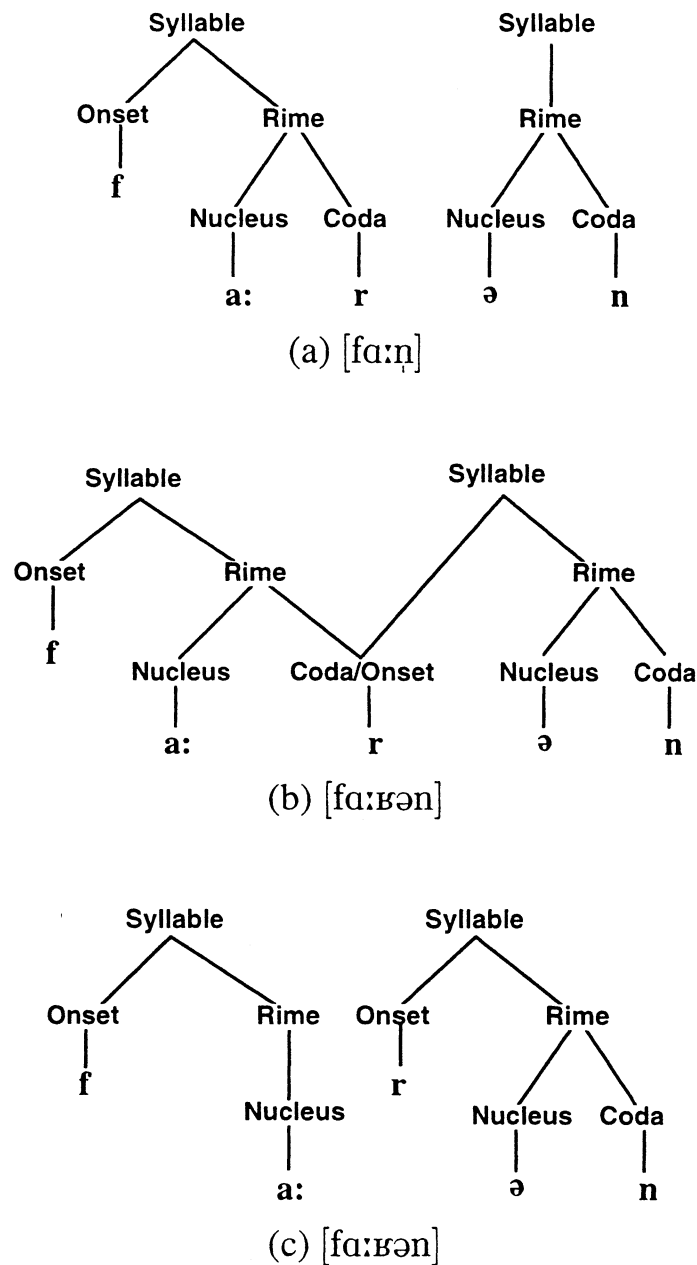


Figure 4: Syllable structures for (a) vocalic and (b) consonantal tokens of the verb *fahren*. If ambisyllabicity is admitted then there are two structural possibilities for consonantal tokens: (b) *r* is at the coda of the first and onset of the second syllable, or (c) *r* is only at the onset of the second syllable.

It is important to consider how this differs from a possible generative account. The necessary phonological abstractions are restricted to syllable structure and the different places in that structure which *r* can take up. Phonetic exponency statements interpret *r* differently depending on its place in structure. The actual patterns which are observed in utterance are the result of the temporal combination of *r*-correlates with those of other objects together with certain articulatory-aerodynamic factors. This allows us to account for a variety of surface phonetics without the need for processes at the phonological level which derive the different patterns from a single base phonetic form.

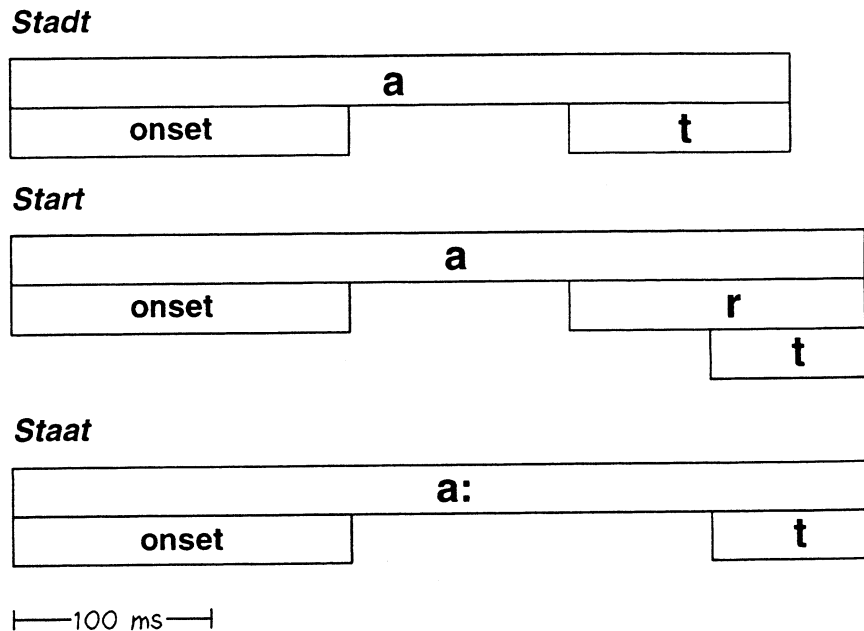


Figure 5: Temporal organization for synthetic versions of the short quantity syllables *Stadt* (“town”) and *Start* (“start”) and the long quantity syllable *Staat* (“country”). The drawing is to scale.

4 Modelling the phonetics

An interesting and challenging method of verifying the analysis presented in the previous section is to use it to drive a speech synthesizer and thus produce acoustic output. The computational implementation described here has much in common with YorkTalk (Coleman 1992; Local 1992; Local and Ogden 1997) and IPOX (Dirksen and Coleman 1997), both nonsegmental declarative attempts at driving speech synthesizers. This applies to the strict division between phonological structure and phonetic exponency and the way in which phonological structure is given a phonetic interpretation. The model outlined here produces control signals to drive an implementation in C of the Klatt (1980) formant synthesizer.

The phonological structure implemented is essentially that illustrated in Figure 4. The phonetic interpretation of the structure begins by giving each node in the phonological structure a start and end time. The temporal extent of each node encompasses the time span of all daughter nodes. Once temporal information has been assigned, the phonetic exponents of each node are laid down. Timing in the phonetic exponency statements carried out relative to the structural starts and ends, and not in absolute terms. The temporal extent of the phonetic correlates of a particular phonological object are the product of the interaction of the timing assigned to the phonological structure and that encoded in the exponency statements. Differences in the length of consonantal articulations following short and long quantity vowels illustrates one advantage of this separation of temporal information. In long quantity syllables the amount of time assigned to the coda is less than that assigned to the coda of a short quantity syllable. The temporal extent of the phonetic exponents of a coda object in a short quantity syllable is then automatically greater than it is in a long quantity syllable, without this being part of the phonetic exponency statement itself.

Figure 5 shows the temporal make-up of synthetic versions of the monosyllabic words *Stadt* (“town”), *Start* (“start”) and *Staat* (“country”). The drawings are to scale and each block represents the time span covered by nodes in the phonological structure of each syllable. Note that

Table 1: Exponency statement to calculate times and values of F1 and F2 for **r** at coda. *start* and *end* refer to the start and end times of the relevant node in the syllable structure; *length* refers to the time span, i.e. *end* – *start*. *target* refers to the formant value of the vowel. *Fx* refers to values of F1 or F2.

Time [ms]	Value [Hz]
<i>start</i> – <i>length</i>	<i>target</i>
<i>start</i>	$Fx + 0.5 \times (target - Fx)$
<i>end</i> – 20	$Fx + 0.5 \times (target - Fx)$
<i>end</i> + 20	<i>target</i>

this is *not* the same as the temporal extent of the phonetic correlates of the objects at each node, which will become clear when we look at part of the exponency for coda-**r** below. The words *Stadt* and *Start* are short quantity, in 5(c) long quantity. The duration of the short quantity syllable in *Start* is greater than *Stadt* due to the increased complexity of the coda. The duration of the coda in both *Stadt* and *Start* is greater than in the long quantity syllable *Staat*. At present the extra duration is assigned to the head of the coda, giving rise to a longer plosive closure in *Stadt* than *Start* and *Staat*. Of greatest interest is the surface similarity exhibited by *Start* and *Staat* despite structural and temporal differences in their make-up. The vocalic portions resulting from each have marginally different durations (the vocalic portion of *Start* is 4 ms shorter).

Part of the phonetic exponency – calculation of times and values for F1 and F2 – for **r** at coda are shown in Table 1. In the left hand column are points in time relative to the structural times (*start*, *end*). So, for instance, for coda **r** *end*+20 refers to a point 20 ms after the end time assigned to the coda node containing **r**. It is now clear how the temporal extent of the phonetic correlates of an object at a particular place in structure differ from the time span assigned to the node itself. The time span of the coda node containing **r** in Figure 5 is 148 ms, but the temporal extent of the phonetic correlates of **r** at this place in structure begin much earlier. The righthand column in Table 1 calculates values for F1 and F2. The qualitative combination of the phonetics of the vowel (*target*) and those of **r** is modelled using a simple locus equation. Values between the calculated points are arrived at using a cosine interpolation. At present, the same locus equation and temporal pattern is used for both F1 and F2 and this undoubtedly represents a simplification, but it is nevertheless sufficient to allow important aspects of the observed patterns to be reproduced.

The separation of timing at different levels allows the different monophthongal and diphthongal qualities to be reproduced using the same exponency statements. The duration of a long quantity syllable is longer than a short quantity syllable (see Figure 5). The duration of the coda in the short quantity syllable is longer than in the long quantity syllable. The temporal extent of the phonetics of **r** at the coda of a short quantity syllable is therefore greater in both absolute and relative terms. In addition to this the time span of the coda node containing **r** is used to define the point in time at which formant movements for **r** begin relative to the start of the node itself (*start* – *length*). The combination of these factors means that the qualitative combination of vowel and **r** in a short quantity syllable often begins before the end of the onset, as is the case in *Start* in Figure 5.

The consequences of these timing differences are illustrated in Figure 6. The figure shows sonagrams of the short quantity *lernt* (“learns”) and long quantity *leert* (“empties”). The phonetics of **r** in *lernt* have already begun during the lateral at the beginning of the syllable, the vocalic portion is monophthongal in quality. In *leert* the diphthongal quality of the vocalic por-

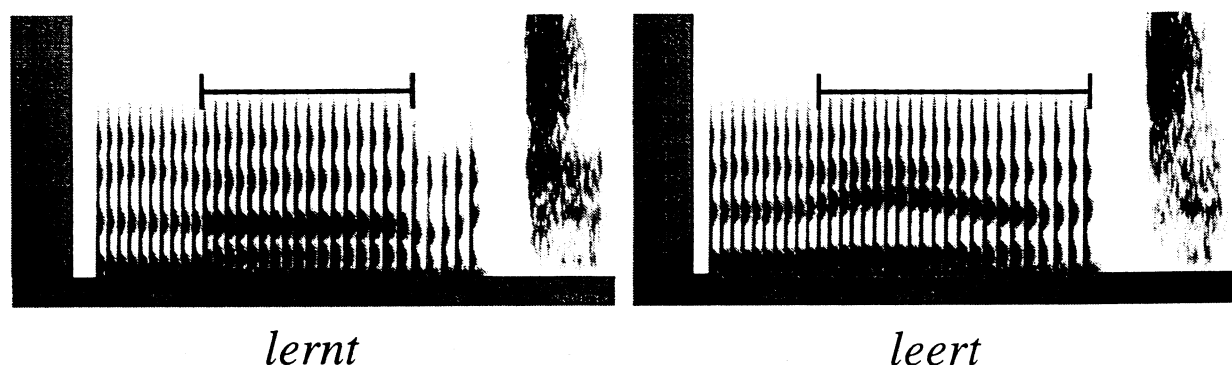


Figure 6: Sonagrams of the short quantity *lernt* ("learns") and the long quantity *leert* ("empties"). The horizontal line approximately delimits the vocalic portion in each case.

Table 2: Examples of synthetic short and long quantity *r* and *r*-less syllables. The words *Bart* and *Dirk* have both been synthesized as short and long quantity syllables.

Word (gloss)	Syllable quantity
<i>Stadt</i> ("town")	short
<i>Staat</i> ("country")	long
<i>Start</i> ("start")	short
<i>bat</i> ("offered")	long
<i>Bart</i> ("beard")	short
<i>Bart</i>	long
<i>Tier</i> ("animal")	long
<i>Kur</i> ("cure")	long
<i>Kür</i> ("free section")	long
<i>Dirk</i> (proper name)	short
<i>Dirk</i>	long
<i>durch</i> ("through")	short
<i>Storch</i> ("stork")	short

tion is clearly visible in movements of F1 and F2. In this case the phonetics of *r* starts some time after the release of the initial lateral and the phonetics of the long quantity vowel are able to 'peek' through for a short time.

Table 2 contains a list of monosyllabic words illustrating the phonetics of coda *r* with various short and long quantity vowels. The words *Bart* ("beard") and *Dirk* (proper name) have been synthesized as both long and short quantity vowels in an attempt to illustrate one of the areas in which distributional differences between individual speakers and dialects can arise. The *r*-less words *Stadt*, *Staat* and *bat* have been included for comparison. Of particular interest here are comparisons of *Staat* and *Start* as well as *bat* and *Bart*. Surface similarity in the temporal organization of these pairs is now joined by surface similarity in the auditory impression, although the vocalic portions in each pair are acoustically and auditorily different.

5 Concluding remarks

This paper has presented a description of the complex consonantal and vocalic patterns associated with Standard German *r*. The description of the vocalic patterns painted a more complex picture than previous analyses and it was claimed that the descriptive oversimplification may

have arisen from the inability to temporally delineate the vocalic correlates of *r* in many contexts. Section 3 provided an account for these patterns in a declarative, Firthian framework and at the same showed how such an approach ultimately affected the way in which the description itself was carried out because different theoretical assumptions played a part in data interpretation. This was particularly the case in those examples which other analyses had considered to be cases of *r*-elision. Finally, in the previous section a computational implementation of certain aspects of the phonetic and phonological analysis was presented which allowed acoustic and auditory inspection of the analytical claims being made.

The ability to produce very similar phonetic patterns on the basis of significant differences in the phonological structure, temporal organization and phonetic exponency raises interesting questions regarding speech production and perception. As was said above, the phonetic identity in pairs such as *Start* and *Staat* is still hypothetical as it could not be tested on the data base material available. However, if it does prove to be the case that speakers produce word pairs which are acoustically identical, finding out whether the productive mechanisms behind the same surface phonetics, which the model predicts, will be difficult to ascertain, and it is not clear at present how this could be done.

References

- Barry, W. J. (1995). Schwa vs. schwa + /r/ in German. *Phonetica* 52, 228–235.
- Bickley, C. A. and K. N. Stevens (1986). Effects of a vocal-tract constriction on the glottal source: Experimental and modelling studies. *Journal of Phonetics* 14, 373–382.
- Bickley, C. A. and K. N. Stevens (1987). Effects of a vocal-tract constriction on the glottal source: Data from voiced consonants. In T. Baer, C. Sasaki, and K. S. Harris (Eds.), *Laryngeal Function in Phonation and Respiration*, pp. 239–253. Boston: College-Hill.
- Bremer, O. (1893). *Deutsche Phonetik*. Leipzig: Breitkopf & Härtel.
- Browman, C. P. and L. M. Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.
- Browman, C. P. and L. M. Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, pp. 341–376. Cambridge: Cambridge University Press.
- Browman, C. P. and L. M. Goldstein (1992). Articulatory phonology: an overview. *Phonetica* 49, 155–180.
- Coleman, J. S. (1992). “Synthesis-by-rule” without segments or rewrite rules. In G. Bailly, C. Benoît, and T. R. Sawallis (Eds.), *Talking Machines: Theories, Models and Designs*, pp. 211–224. Amsterdam: Elsevier.
- Coleman, J. S. (1994). Polysyllabic words in the YorkTalk synthesis system. In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, pp. 293–324. Cambridge: Cambridge University Press.
- Dirksen, A. and J. S. Coleman (1997). All-prosodic speech synthesis. In J. P. H. v. Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Eds.), *Progress in Speech Synthesis*, pp. 91–108. Berlin/Heidelberg/New York/Tokyo: Springer.
- Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*, 127–152.

- Graf, J. and B. Meißner (1996). Neue Untersuchungen zur r-Realisation. *Hallesche Schriften zur Sprechwissenschaft und Phonetik 1*, 68–75.
- Hall, T. A. (1993). The phonology of German /r/. *Phonology 10*, 83–105.
- Henderson, E. J. A. (1949). Prosodies in siamese. *Asia Minor 1*, 189–215.
- IPDS (1994). *The Kiel Corpus of Read Speech*, Volume 1, CD-ROM#1. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- Kelly, J. and J. K. Local (1989). *Doing Phonology*. Manchester: Manchester University Press.
- Klatt, D. H. (1980). Software for a cascade/parallel synthesizer. *Journal of the Acoustical Society of America 67*, 971–995.
- Kohler, K. J. (1990). German. *Journal of the International Phonetic Association 20*(1), 48–50.
- Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen* (2 ed.). Berlin: Erich Schmidt.
- Kohler, K. J., M. Pätzold, and A. P. Simpson (1995). *From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech*. AIPUK 29.
- Lisker, L. and A. S. Abramson (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word 20*, 384–422.
- Local, J. K. (1992). Modeling assimilation in nonsegmental, rule-free synthesis. In G. Docherty and D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pp. 190–223. Cambridge: Cambridge University Press.
- Local, J. K. and R. Ogden (1997). A model of timing for nonsegmental phonological structure. In J. P. H. v. Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Eds.), *Progress in Speech Synthesis*, pp. 109–121. Berlin/Heidelberg/New York/Tokyo: Springer.
- Meinhold, G. (1989). Das problematische [v]. In Slembek (Ed.), *Von Lauten und Leuten. Festschrift für Peter Martens zum 70. Geburtstag (Sprache und Sprechen 21)*, pp. 119–125. Frankfurt am Main: Scriptor Verlag.
- Nooteboom, S. G. (1972). *Production and Perception of Vowel Duration*. Ph. D. thesis, Rijksuniversiteit te Utrecht, Utrecht.
- Scobbie, J. M. (1993). Constraint violation and conflict from the perspective of declarative phonology. *Canadian Journal of Linguistics 38*, 155–169.
- Stevens, K. N. (1987). Interaction between acoustic sources and vocal-tract configurations for consonants. In *Proc. XIth ICPHS*, Volume 3, Tallinn, pp. 385–389.
- Ulbrich, H. (1972). *Instrumentalphonetisch-auditive R-Untersuchungen im Deutschen*. Schriften zur Phonetik und Kommunikationsforschung Nr. 13. Berlin: Akademie Verlag.
- Viëtor, W. (1894). *Elemente der Phonetik des Deutschen, Englischen und Französischen*. Leipzig: Reisland.