# The investigation of connected speech processes. Theory, method, hypotheses and empirical data*

Klaus J. Kohler

*The writing of this paper benefited from fruitful discussions with Benno Peters and Thomas Wesener.

# 1 Overview of past research at IPDS Kiel

The study of phrase-level phonetics has been a focal research area in the Department of Phonetics and Digital Speech Processing at Kiel University for about three decades. The interest in phonetic phenomena above the word in utterances and larger pieces of connected speech took its point of departure in the auditory and descriptive account of segmental assimilation and reduction of German content and function words in text reading, compared with unscripted speech. To begin with, the investigation was carried out within a German Research Council funded project on the phonetic manifestations of speaking styles in Standard North German (DFG Ko 321/1ff) between 1972 and 1977. In the initial stage, it focussed on auditory processing of corpus data, supported by spectrographic analysis, but was then supplemented by lab speech experiments in coarticulation and articulatory control (Kohler 1974, 1976; Gnutzmann 1979; Butcher, Kohler, and Kuenzel 1977).

## 1.1 Taxonomic description of corpus data

The results of the corpus investigations were summarized in a set of ordered generative rules for assimilation and elision to be applied to canonical phonemic strings, including various types of boundaries (Kohler 1974), in Standard North German. The function of these rules was the systematic enumeration of all the phrase-level phonetic changes that were found in the corpora, with reference to canonical word forms. The rules were not intended to represent a model of speech production, but the reductions they captured were interpreted in articulatory terms. The rules covered

(1) schwa elision as the result of articulatory economy in unstressed syllables of rhythmic feet: *e<u>ben</u>* ‘even’ [bən] > [bn]

(2) economy of tongue tip articulation: perseverative assimilation of apical to preceding labial or dorsal articulator in nasals or unreleased plosives after nasals or plosives, within words: *e<u>ben</u>* ‘even’ [bn] > [bm]

(3) economy of tongue tip articulation: perseverative assimilation of apical to preceding labial or dorsal articulator in nasals after fricatives, within words: *rau<u>chen</u>* ‘rauchen’ [xn] > [xŋ]

(4) economy of tongue tip articulation: anticipatory assimilation, between plosives and nasals, of apical to following labial or dorsal articulator, also accross morpheme and word boundaries: *in Berlin* [nb] > [mb]

(5) simplification of complex consonant articulations: /t/ deletion inside triple clusters, also across word boundaries: *einst stritten* 'were once disputing' [stʃ] > [sʃ]

(6) assimilation of narrow-groved fricatives to following wide-groved ones across word boundaries, to reduce sequential complexity: *einst stritten* [sʃ] > [ʃʃ]

(7) change of plosives to fricatives/approximants in inter-sonorant position in stop target undershoot: *wirklich* 'really'; [iːɐkl] > [iːɐxl]; *ich habe* 'I have' [aːbə] > [aːβ̞ə]

(8) lenition of voiceless fortis obstruents to (voiced) lenis obstruents in sonorant environment if the entire chain is in the remiss of a rhythmic foot, due to reduced degree and duration of stricture formation, also across word boundaries: *das muß ich máchen* 'I must dó this' [mʊsɪç] > [mʊz̥/zɪç]

(9) complete devoicing of voiced obstruents after voiceless obstruents, due to aerodynamic and intra-oral pressure conditions in obstruents, also across word boundaries: *setzt sich* 'sits down' [tsz] > [tss]

(10) partial devoicing of voiced approximants after voiceless fortis obstruents, due to aerodynamic and intra-oral pressure conditions in approximants after obstruents, also across word boundaries: *das Wasser* 'the water' [sv] > [sv̥]

(11) anticipatory assimilation of lenis plosives to following consonant nasality, due to early velum lowering in short stop closures, inside words and fused phrases: *eben* [bm] > [mm]

(12) perseverative assimilation of lenis plosives to preceding consonant nasality, due to late velum raising in short stop closures, also across word boundaries: *zum Beispiel* 'for example' [mb] > [mm]

(13) geminate reduction inside words (after schwa deletion) and across word boundaries, especially in the remiss of the rhythmic foot, due to stricture weakening: *einstststritten* [ʃʃ] > [ʃ], *setzt sich* [tss] > [ts], *eben* [mm] > [m], *zum Beispiel* [mm] > [m]

Except for the devoicing rules, all the rules are optional. In addition to the contextual and accentual conditions mentioned, their application depends on speaking style and speech rate, and their combined force results in a large variability of phrase-level phonetic word forms, which are the more reduced the more casual the speaking style and the greater the grammatical and prosodic cohesion between words and phrases.

## 1.2 Experimental data and phonetic explanation

The results of the corpus study, incorporated in these rules, did not go beyond the taxonomic stage. The limited corpus and the restricted tools for data base analysis, available at the time, provided a corpus-based description that could not give an answer as to whether certain reduced forms were missing by chance, or whether they were precluded due to systematic constraints in the physiological and articulatory requirements of human speakers, on the one hand, and in the socio-cultural modulation of these biological and physical conditions, on the other. The elucidation of the former constraints necessitated experimental investigation. A series of experiments were devised to deal with some of the physical parameters of reduction phenomena in systematically varied utterances of sentence length, aiming to explain some of the regularities previously found in the corpus investigation. The outcome was a theory of plosive articulation and of coarticulation in intervocalic obstruents in conneted speech as a basis for explaining place assimilation and nasalization in plosives, and lenition in plosives and fricatives. In this connection, the fortis/lenis distinction was also given a new substance-based definition as an important prerequisite to the explanation of plosive-related phrase-level phenomena (Kohler 1976; Butcher, Kohler, and Kuenzel 1977).

In the instrumental investigations, the timings of the various articulatory and phonatory parameters as well as their synchronizations became the focus of attention. This led to a reformulation, and at the same time explanation, of the phrase-level reduction phenomena (Kohler 1990). Changing strings of static phonetic symbols in the sequence of rules was not

able to capture the temporal phonetic processes, because unitary articulatory patterns were decomposed into a series of segmental steps. For instance, the change from [kʰən] to [kŋ] is achieved by gestural reorganization which eliminates an opening-closing of two articulators – tongue dorsum and tongue tip, replaces it by a long oral closure of the dorsum and transfers the stop release to the velum. This new coordination of articulatory structures and their movements in time results in the simultaneous deletion of aspiration, schwa and apical closure, which is misrepresented by the segmental sequence [kʰən] **>** [kʰn] **>** [kn] **>** [kŋ] provided by the rules.

The EPG investigation of place of articulation in plosives across word boundaries (Kohler 1976) provided the following evidence:

(a) Only apicals are changed, i.e. the sequences LABIAL # DORSAL and DORSAL # LABIAL stay unaltered, except for temporal overlap.

(b) Apicals in APICAL # LABIAL/DORSAL may be assimilated, but the sequences LABIAL/DORSAL # APICAL are only coarticulated, i.e. temporally overlapped.

Furthermore, the corpus data allow the generalizations that

(c) apicals in APICAL + ə + LABIAL NASAL and in LABIAL/ DORSAL + ə + APICAL NASAL are assimilated as well, and

(d) only stops and nasals are assimilated, fricatives are only coarticulated.

Three questions have to be answered as regards these assimilation and coarticulation patterns:

- Why can apical, rather than labial or dorsal gestures be eliminated?
- Why is this process restricted to stops and nasals?
- Why does it only happen *before* labials and dorsals across word boundaries and not to stops *after* labials and dorsals?

Some of the answers may be found in speech production, others in acoustics and perception. Speech articulation makes continuous use of

dorsum and of lip movements in temporally varying ways, for vowels and consonants likewise. Superimposed on these continuously ongoing basic articulatory processes are additional apical tongue movements, limited, in the main, to consonants. Tongue dorsum and lip gestures, in close conjunction with jaw movements, constitute large, global processes, associated with, e.g., the large extrinsic tongue muscles. Tongue apex articulations, on the other hand, require more finely controlled and more precisely tuned, and therefore also more costly, adjustments, linked with the smaller intrinsic tongue muscles. When the perceptual differentiation between assimilated and non-assimilated apical gestures is small, the fine muscular control may be dispensed with under certain situational conditions – relaxed speech with reduced energy expenditure per unit time (Lindblom 1983), or higher speech rate with less time for finer control. Thus, it is the additional, occasional and special apical gestures, requiring greater precision of timing and muscular coordination than the basic, continuously present labial and dorsal movements, that are weakened and eliminated. This reference to an economy of speech effort within the physiological conditions of the vocal apparatus can explain why only sequences containing apicals are affected by these assimilation processes, and why they do not result in the preservation of apicals.

Moreover, these assimilations do not affect word and syllable initial apicals because the initial position has a particularly high signalling value for a listener and thus demands greater articulatory precision from a speaker than the final position. There are stronger and more numerous cues to identify place of articulation, especially for stops, which are differentiated by spectral characteristics of the burst and (in the case of voiceless ones) of aspiration, over and above the differences in formant transitions.

Furthermore, fricatives are normally not assimilated in their active articulators (lips, tongue apex or dorsum), because, due to different mechanisms in the generation of turbulence, they are acoustically and auditorily far more distinct than nasals and unreleased plosives with regard to place cues so that their articulatory reduction would be too salient for a listener and is, therefore, not produced by a speaker.

The experimental investigation into the production of plosives in various phonetic contexts has shown that lenis plosives in vowel + plosive sequences have short closure (but long preceding vowel) durations, and that for fortis plosives it is the reverse. This finding may be connected with the corpus observation that nasal assimilations of plosives before nasal conso-

nants are tied to lenis (cp. *lagen* [gŋ] > [ŋ] versus *Laken* [kŋ]). If in relaxed or fast speech the raising of the velum for a stop before an immediately following nasal consonant is curtailed because the trajectories are shortened, the short stop, resulting from oral + velic occlusions, may be decreased still further and eliminated altogether. This does not happen in fortis stops outside function words, because there the closure, and therefore the temporal safety zone, is much longer. What may happen, however, is lenition and voicing through shortening of the closure duration.

In unstressed function words, articulatory movements are weakened even more, resulting in reduced closure duration and less firm contacts for voiceless fortis stops so that passive voicing may continue through the occlusion. Thus *mit dem* 'with the' [mɪpm] may change to [mɪbm], and this lenis stop may in turn be subject to nasal assimilation. The same applies to content words that have become devoid of meaning in set phrases, such as adjectives in greetings, e.g. *guten Tag* 'good day' [gʊdn]. The fact that [mɪpm] may develop to [mɪm], or [gʊdn] to [gʊn], whereas *Lappen* 'cloth' [lapm] always retains a stop closure, can therefore find its explanation in a greater reduction of articulatory energy expenditure under lack of stress, and in the higher probability of lenis stops, rather than the much longer fortis ones (under equal contextual conditions), losing their velic closure before nasals.

## 1.3 Theory of phrase-level phonetics in speech communication

By linking corpus data of connected speech with experimental results in the production of plosives it has been possible to relate key aspects of place assimilation, nasalization and lenition found in corpus material to articulatory time courses of speech production. This has in turn enabled the explanation of some connected speech phenomena with reference to language-external conditions in the speech generating apparatus. This first step towards an explanatory approach has subsequently been developed into a theory of phrase-level phonetics, combining a system of conditions on speech production in communication with a new concept of basic speech gestures and of the programme operating on them.

### 1.3.1 Conditions on speech production in communication

Phrase-level speech phenomena are controlled by the following system of conditions in speech communication.

(1) They are grounded in speech production and in the principle of ***goal-oriented motor economy in the speaker***: the vocal apparatus is not put to more effort than is necessary to reach a communicative goal.

(2) The speaker's motor control of phrase-level speech output is checked by the need to maintain ***sufficient linguistic distinctivity for the listener*** (Lindblom's (1990) H&H theory). This is achieved by three factors:

(2.1) The likelihood of reductions is inversely related to ***the degree of their perceptual salience*** in comparison with unreduced forms. What is not a very noticeable change for a listener will be produced more easily by a speaker to yield to the principle of motor economy. This can be considered the theoretical basis for the empirical findings that apical fricatives, released apical plosives and syllable-initial, as well as word-initial, apical nasals and plosives are not assimilated, progressively or regressively, as to their places of articulation, whereas final nasals and unreleased plosives are. Perceptual conditions change in sequences of unstressed syllables, including function words, where place-assimilated nasals initially in a syllable, or even in a word, are very short, consequently reduced in their perceptual salience, and thus do occur, as in *hervorragende Idee* 'excellent idea' [aːŋŋə], *vorschlagen, daß* 'suggest that' [aːŋŋas] (examples from the *Kiel Corpus of Spontaneous Speech*, g094a021 and g145a010; see also (3.2)).

(2.2) There are also ***two situational constraints***, which Kohler (1979) introduced into phrase-level speech production theory on the basis of concepts from Speech Detection Theory (predating Lindblom's H&H theory).

(2.2.1) On the one hand, there are ***the a priori probabilities of communicative failure***, different for different communicative settings, e.g. lower for two-way face-to-face spontaneous interaction, higher in one-way formal lectures with reduced signal/noise ratios, possibly

still higher in a complex style of scripted speech. Thus the different communicative situations and the speaking styles attributable to them require different degrees of articulatory precision based on varying amounts of effort on the part of the speaker. Speakers have intuitive knowledge of these requirements and usually adjust their speaking more or less to them. This in turn means that different speaking situations are correlated with different degrees of reduction or reinforcement in phrase-level speech, and the statistics of reduction phenomena across different speaking styles can be expected to show different distributions.

(2.2.2) But over and above these a priori probabilities, anchored in the communicative situation, there are also different ***cost assessments for communicative failure***, which speakers associate with the particular communicative situation. Their rating of the cost of failure in a speech act may vary a good deal from speaker to speaker and may even change within an interaction in accordance with its function and the information to be conveyed. This cost aspect can interfere with the covariance of type of speaking style and phrase-level phonetic phenomena.

(3) In controlling listener-oriented motor economy at the phrase level, the speaker also takes ***linguistic patterns*** into account.

(3.1) There is a need for ***paradigmatic differentiation of linguistic units***. It influences the degree of reduction according to word class – e.g. function vs. content word – and morphological category within function words. German *ihr* [iːɐ] may be personal pronoun, 2nd person plural, nominative 'you', or personal pronoun, 3rd person singular feminine, dative 'to her', or possessive pronoun, 3rd person singular feminine (or 3rd person plural or 2nd person polite), uninflected 'her' ('their'/'your'). It may be reduced [ɐ] at a low stylistic level if it is the 2nd person subject pronoun in enclitic position *(habt ihr das gemacht?* 'have you done it?'); this is less likely in the 3rd person indirect object pronoun, and impossible in the possessive pronoun (*Er hat ihr geholfen.* 'He has helped her.'; *ihr Wagen* 'her car'). In the latter case, the syntactic parsing also gives *ihr* the status of a proclitic as against enclitics in the other examples. So the factor of

syntagmatic structure also intervenes in the reduction of these pronouns.

(3.2) There is also a need for ***syntagmatic structure in linguistic messages*** at a hierarchy of levels from syllables to words to morphological and syntactic constructions to semantic organization, and to prosodic grouping by accent, intonation and phrase boundaries. The prosodic features may support, or cut across, any of the former syntagmatic elements. These groupings are characterised by ***internal cohesion and junctural separation at the boundaries*** (Kohler 1983), signalled by segmental and prosodic indices. Internal cohesion raises the probability of phonetic fusion inside the various syntagmas, whereas their boundaries have a high probability of being marked by phonetic separators (Kohler 1991). For instance, in German

- post-stress enclitics in rhythmic foot structures are reduced more than pre-stress proclitics (*hat er geschrieben?* 'has he written?' [hadɐ] vs. *er hat* geschrieben 'he has written' [ɛɐhat])
- modal particles are reduced when they form enclitics to verbs (+ pronouns), but not after nouns (*woher kennst du denn/denn du meinen Chef?* [n] vs. *woher kennst du meinen Chef denn? wo kennst du meinen Chef denn her?* [dɛn] 'how do you know my boss?')
- phrase-finally there is less reduction than phrase- medially (*Wo kommt er denn her?* [n] 'where is he from?' vs. *Kommt er denn?* [dɛn] 'is he coming?'(at the same stylistic level))
- left-to-right place assimilation across word boundaries depends on the absence of an intervening prosodic boundary, as in *vorschlagen, daß* 'suggest that' [aːŋŋas] (see also (2.1), where the syntactic boundary between main and subordinate clause is not marked prosodically).

(3.3) Moreover, the cohesive units are flexible in extension; e.g., sequences of words may fuse to new lexical items, as in German *zum* (< *zu dem*) 'to the', or French *du*, *au* (Kohler 1990, 1998). The importance of linguistic patterning over and above goal-oriented motor economy at

the phrase-level makes it necessary to consider the category of ***lexicalization*** beside synchronic statistical variation of phonetic forms.

This may be illustrated by the German ordinal number 'seventh'. It is either *siebten*, derived from the monosyllabic stem of the cardinal number *sieben*, thus making all ordinals up to 'twelfth' bisyllabic; or it is derived from the reduced form of the cardinal number *sieben* [ziːm], which is also monosyllabic. Today, the canonical form of *sieben* is [ziːbm], with reinforcements to [ziːbən] and [ziːbn], and with further reduction to [ziːm]. In *siebenten*, however, the canonical form is [ziːmtən], with any trisyllabic realization sounding elaborated and orthography-driven. There are thus two lexical forms of this ordinal number, and each may follow a course of articulatory reduction:

| I(a) | I(b) | II(a) | II(b) |
|---|---|---|---|
| [ziːmtən] | [ziːmpm] | [ziːptən] | [ziːpːm] |
| [ziːmtn] | [ziːm$\widehat{\text{ʔ}}$pm] | [ziːptn] | [ziːʔ$\widehat{\text{p}}$ːm] |
| [ziːm$\widehat{\text{ʔ}}$tn] | [ziːmm̰m] | [ziːp$\widehat{\text{ʔ}}$tn] | |
| [ziːmn̰n] | | | |

(4) Finally, the speaker follows ***social conventions***. The language community judges the social and stylistic acceptability of phrase-level phonetic variation, as it results from the principle of motor economy on the part of the speaker, and of sufficient auditory differentiation for the listener in different communicative settings. Speakers adhere to these social speech codes and adjust their phonetic output to national, regional and social dialects as well as to speaking style conventions, e.g. along a scale from formal to casual, in their communicative environment.

Phrase-level reduction varies a great deal from dialect to dialect and among regional or social accents of the same language. Whereas the simplification of plosive + schwa + nasal opening-closing gestures has affected the opening part in Standard North German (*legen* [gən] > [gŋ]), it is the closing part in South German dialects ([g$\overline{\text{ən}}$] > [gə]). Similarly, languages differ in the way they exploit the universal directions of speech effort reduction in the same types of phrase-level phonetic structures. Thus in English plosive + schwa + nasal opening-closing gestures the levelling of the opening does

not normally extend to articulator adjustment (*organ* [gən] > [gn]), at least not in the varieties (and their stylistic levels) that have been described most extensively. Moreover, in **nasal** + plosive + schwa + nasal structures the vocoid opening is not eliminated at all in RP and other varieties of Southern British (*London* [ndən], *Clinton* [nt$^{h}$ən]), as far as research has established up to now. In American English, on the other hand, the same reduction and glottalization occurs in the fortis plosive type of this contextual structure as in German ([ntən] > [nn̰n]). So in all these cases, the general articulatory propensity is checked in different ways by social conventions in the different speech communities.

(5) Phrase-level phonetics is thus shaped by the relation between

- ***the speaker's tendency to reduce effort***
- ***the listener's demands for distinctivity***
- ***the distinctivity requirements set by the communicative situation***
- ***linguistic patterns***
- ***social conventions***.

So the theory of phrase-level phonetics is part of a ***theory of speech communication***.

### 1.3.2 Phrase-level reduction scales, speech production units and speech gesture control

The effect of these five conditioning factors on phrase-level phonetic output cannot be deterministic, but leads to statistically variable reduction patterns, some of which have very high, others quite low frequencies. Different degrees of reduction in specific types of phonetic sequences that result from the five factors operating with variable strength can be arranged along an articulatory scale from canonical phonological form to extreme simplification, as in the German preposition + definite article structure *mit dem*

| I. | II. | III. |
|---|---|---|
| [mɪtʰd̥eːm] | [mɪt͡pːm] | [mɪpm] |
| [mɪtd̥em] | [mɪpːm] | [mɪb̥m]] |
| [mɪtd̥əm] | [mɪʔpm] | [mɪbm] |
| [mɪtːm] | | [mɪmm] |
| | | [mm] |

As was pointed out in 1.2, the linear segmental representation of these forms hides the temporal patterns of coordinated articulatory parameters and creates the false picture of discrete segmental changes. The segmental phonetic transcriptions must therefore be translated into the timing of 'supra-segmental' articulatory patterns, and it is these superordinated speech gestures that are changed in respect of the synchronization and the composition of their articulatory parameters. These more global units of speech production are defined by the ***opening-closing gestures of the vocal tract***: articulated speech can be viewed as being composed of sequences of these.

There are three such gestures in the above example, and it is the second and third that exhibit varying degrees of reduction. The opening between stops is eliminated, which results in the absence of an audible release, and in the case of identical articulators for the two stops, produces a single occlusion; in the case of different articulators there is temporally sliding coarticulation, i.e. the second articulator forms a closure at a variable point during the hold of the first.

The third opening-closing gesture has three dimensions: I. its vertical opening, II. the chaining of the apical and labial articulators, which make up the gesture, and III. the synchronization of the timings of oral and velic closures. In block I, the vertical dimension varies from a long highish front vocoid opening via a mid-central vocoid opening to no vocoid opening at all; the apical and labial articulators as well as the oral and velic closures remain well separated. In block II, the chaining of the two articulators takes place, from overlapping coarticulation to a complete reorganization of the whole opening-closing gesture by eliminating the apical movement. In block III, the timing of the nasal release of the stop closure is successively advanced, thus shortening the all-round closure of the vocal tract and creating the pressure and aerodynamic conditions for passive voicing, then eliminating the momentary stoppage of air altogether, and finally reducing the duration of the oral closure.

By taking opening-closing gestures as the basic constituents of articulated speech and by defining them in their vertical and temporal dimensions with regard to the component articulators and cavities (oral, nasal, laryngeal) of the vocal tract we can transcend the atomic segmental switches of the alphabetic transcriptions and look at the production, modification and reorganization of whole syllable-size units[†]. It is the whole opening-closing gesture that is affected at the phrase level, not the isolated segments that descriptive phonetics may excerpt from this more global unit. When articulatory and physiological constraints of the apical articulator versus the labial and dorsal ones and of the cavity components of the vocal tract (see 1.1/2) are added to this basic concept of opening-closing gestures, and when it is then subsumed under the assumptions (2) to (5) of a theory of phrase-level phonetics in speech communication, as outlined above, all the phonetic forms of *mit dem*, recorded in connected speech, can be derived as different degrees of opening-closing gesture levelling, and explained with reference to a net of communicative conditions.

In the chaining of opening-closing gestures, the end of one can at the same time be the beginning of the next. Internally permissible levelling processes can then be extended across their juncture if the conditions apply, as in *drängt ein bißchen* 'is somewhat urgent' [ɛŋtən] > [ɛŋkŋ] (from the *Kiel Corpus of Spontaneous Speech*, g077a000). Here the lexicalised weak form [ən] of the indefinite article *ein* is probably selected (see (3.3)), and, combined with the elimination of a vocoid opening, the left-to-right assimilation of apical to dorsal is carried through into the following gesture, because the nasally released plosive has not got strong perceptual salience as to place of articulation. However, this coordination of successive gestures is not mechanical, but must be planned, as is shown in *mit bunten Papierschlangen* 'with coloured paper streamers' [ʊntnp] > [ʊmpmp], where the labial trigger occurs at the beginning of a later gesture and operates right-to-left into the preceding two gestures.

Under the view advocated here, it is not necessary to specify all the possible types of assimilation and reduction for individual segmental sequences, but quite general instructions, as given above in the discussion of German *mit dem*, suffice to trigger the application of a reduction program

[†] The term 'syllable' refers to the phonological category, the term 'gesture' to the unit of speech production. The two are not congruent in their syntagmatic reference; e.g. German *stopfst* '(you) darn' is one syllable, but two gestures. 'syllable-size' points to 'larger than sound segment'.

to specific segmental sequences that fall into various types of opening-closing gestures. Speakers do not set out to alter an underlying phonemic chain in a series of steps before handing it over to the articulators for execution, as would be suggested by a generative phonology, nor do they select all the different forms from a lexicon, and the changes are not simply caused by peripheral vocal tract constraints in temporal sliding and dynamic modification between neighbouring segmental gestures taken from a gestural score either (Browman and Goldstein 1992). It may be assumed that speakers start from canonical forms, stored in a mental lexicon, as representations of sequences of opening-closing gestures with temporal information. The canonical forms that are entered into such a lexicon are a question of empirical research. At present we can only say that they do not coincide with the most elaborated forms, because they may themselves be reinforcements from canonical forms, as is probably the case for German words ending in *-en*.

In the application of a reduction program to canonical forms, retrieved from a mental lexicon, a speaker has to set a value to the generation system that tells it what degree of reduction it is to achieve. This value is determined by the five conditions set out in (1) to (5). So what the system requires is a ***reduction coefficient*** which would operate at a high processing level before actual articulatory execution. This concept is no more than an abstract theoretical construct, because it cannot be quantified at present, but in order to come to grips with the full diversity of phrase-level speech communication such a concept must be given further elaboration.

### 1.3.3 Phrase-level Phonetics and Articulatory Phonology

The concept of speech gestures outlined here as part of a theory of phrase-level phonetics in speech communication differs in important points from that of Articulatory Phonology (Browman and Goldstein 1992):

- The theory of phrase-level phonetics moves a good deal further away from a linear segmental phonemic framework because the primary gestures are considered to be syllable-size units and articulatory variability is construed as a reorganization of these global dynamic structures according to internal and external conditions. Segment-size units are secondary and result from segmentation of the global gestures into which they are integrated.

- This contrasts with the gestural score in Articulatory Phonology, which is based on successive phonological segments as the primary elements whose coordinated gestural parameter specifications are temporally and dynamically variable in concatenation.

- Syllable-size gestures incorporate segments and long componential features, as units in their own right. Articulatory Phonology only recognises segments, feature spreading is a consequence of segmental gestural sequencing.

- In the theory of phrase-level phonetics, gestural interaction applies to global structures of flexible extension and with a high degree of internal cohesion, in Articulatory Phonology the interaction is local between juxtapositioned segments. This basic approach has not been changed by the introduction of the concepts of a phase window and of prosodic boundary or $\pi$-gestures in the development of Articulatory Phonology by Byrd (1996) and Byrd et al. (2000).

- Components of opening-closing gestures may disappear or be changed to others; in Articulatory Phonology gestures that are incorporated in the gestural score cannot disappear nor be changed to other gestures.

- Syllable-size gestures are not articulations that are deaf to the auditory consequences, as is the case in Articulatory Phonology.

- Syllable-size gestures are embedded into communicative functions which determine their realization, whereas Articulatory Phonology does not incorporate the functional aspect and treats gestures as mechanical processes without cognitive links.

- Articulatory Phonology has always worked with highly stylised scripted spech and so far not attempted to deal with spontaneous speech processes. At the present stage of the theory, it would not be able to cope with the extremely variable degrees and extents of phrase-level gesture modification.

# 2 A new German Research Council project

By the middle of the 1990s, the theory of phrase-level phonetics in speech communication had advanced quite considerably, and in parallel to it, IPDS Kiel had begun to collect and phonetically annotate two large data bases of connected speech of Standard North German in computer-accesible form (*The Kiel Corpus of Read/Spontaneous Speech*, IPDS 1994ff), within the PHONDAT and VERBMOBIL projects (Kohler, Pätzold, and Simpson 1995). It was therefore possible to formulate hypotheses, on the basis of the theoretical discussions and the limited early data analysis, about the occurrence of phrase-level phonetic patterns in large amounts of scripted and unscripted speech, and to look at their frequency distributions. This return to the question of corpus analysis with more copious material, better processing facilities and hypotheses based on a theory of phrase-level phonetics led to the German Research Council funded project 'Sound patterns of spontaneous German – symbolic structures and temporal processes' (DFG Ko 331/22-1f) from 1998 onwards.

## 2.1 Methodological considerations

Work was able to start with a fully developed speech processing platform, comprising an annotation framework, a speech analysis software, a data bank generation system and search routines for phonetic data retrieval. The requirements for an annotation procedure are a structured symbol inventory and systematic conventions for its use, including componential markers by the side of segmental ones, as part of a symbol-to-signal processing platform, nowadays within an electronic environment. If these are complemented by standardized canonical word representations and categories for modifying these in relation to the actual pronunciations to be labelled – deletions, insertions, replacements of canonical symbols – the labelling becomes a powerful tool for processing large speech data bases and integrating them into a structured data bank. Such a data bank then makes it possible to sort and access speech data via their symbolic categorizations, to perform classifications and statistical evaluations on the latter, and to analyse classes of speech records that have been selected for specific phonetic questions through their symbolic processing. So, for example, classes can be formed automatically of all the instances of specific manner or place

or phonation modifications or deletions in symbolically specified contexts. The link of the static symbolic abstractions with the dynamic speech signals in these classes not only allows quick access to, and analysis of, the latter but also the setting up and testing of hypotheses about the variability of certain global patterns of articulatory dynamics in synchronic data of a particular language.

Such a research environment is provided by *xassp* (***A***dvanced ***S***peech ***S***ignal ***P***rocessor under the ***X*** Window System, IPDS 1997b) and by *Kiel-Dat* (Kiel data bank utilities, Pätzold 1997) as well as by the CD-ROMs of the *Kiel Corpus of Read/Spontaneous Speech* for German (IPDS 1994, 1995, 1996, 1997a). The labelling framework is basically linear segmental phonemic, with reference to the phonological elements of German, in modified SAMPA notation, but includes symbols for the glottal stop and for plosive release (including aspiration), as well as for componential elements, such as nasalization (**-˜**), glottalization (**-q**) and a general marker of secondary articulation (**-MA**) (Kohler, Pätzold, and Simpson 1995). In actual labelling of signal files by trained phoneticians, the broad phonetic segmental categories are linked to segments of the speech wave, i.e. signal durations are attributed to them. On the other hand, the componential category additions for the symbolization of vowel nasalization (in the case of nasal consonant deletion), of glottalization instead of, or in addition to, glottal stops or plosives, and of secondary articulation residues (in the case of segment deletions) are associated with points in time, but are not given durations, i.e. they only refer to places in signals where the parameters symbolized by them occur, with unspecified extensions to their left and/or right (Helgason and Kohler 1996; Kohler 1994, 1998; Rodgers, Helgason, and Kohler 1997). Further narrow phonetic features, e.g. specific places on the palate for tongue dorsum articulations, are ignored (e.g. [kgŋ] stand for dorsal, not velar). Segmental uncertainties in labelling are successfully resolved (a) by applying a broad, rather than a narrow, symbolization system, and (b) by the use of **-MA** and other componential markers, as they avoid the need for segmentation and segment categorization, but at the same time supply information on linguistically relevant 'suprasegmental' traces, and also allow a systematic renewal of connection with the speech files.

The label files generated with this processing platform do not have the status of descriptive accounts of the phonetics of connected speech in German, but are heuristic devices for systematically retrieving data from large databases for further data analysis, statistical evaluation and inter-

pretation of a large spectrum of phonetic questions. For this goal, the construct of canonical forms is invaluable, and so are the constructs of segmental labelling, segment modification, segment deletion, segment insertion, as well as componential addition. They assure the grouping of extremely variable phonetic data round lexical items for comparative and context-sensitive symbolic processing, for easy and systematic access to the associated speech files and for their analysis. Labelling is thus symbolic preprocessing of speech, facilitating repeated controlled speech wave examination.

Relying on these tools and constructs the question of variable articulatory dynamics in opening and closing movements of German connected speech was investigated with reference to the observable phenomena of vowel and consonant related glottal stop and glottalization, vowel modification and deletion, place assimilation, lenition, nasalization, lateral and geminate reduction, and, more specifically, the manifestation of function words, which contain all these phonetic aspects.

## 2.2 Hypotheses on speech gesture reduction in phrase-level phonetics, related to canonical lexicon entries in German

The phonetic areas to be searched in the data bank were grouped round 10 hypotheses, derived from earlier empirical work and from theoretical discussion. Starting from the hypotheses, search scripts were written for context-sensitive data retrieval in the selected phonetic fields listed above. Statistical patterns were then established across the various data sets to provide a comprehensive descriptive account of the variability of phrase-level reduction patterns and to relate the empirical findings back to the hypotheses and the theory.

**Hypothesis 1 – condition of speaking style**
Ceteris paribus all phrase-level phonetic phenomena are more frequent in unscripted than in scripted speech.

**Hypothesis 2 – condition of stress**
Ceteris paribus all phrase-level phonetic phenomena are more frequent under lack of word stress and sentence accent and are most common in function words.

**Hypothesis 3 – condition of syntagmatic cohesion**
The cohesion of syntagmatic units supports, or interferes with, the occurrence of the phrase-level phonetic phenomena, depending on whether they are placed inside these units or span their initial boundaries, respectively. Such cohesive units are phonological/phonetic (syllables/opening-closing gestures) or morpho-syntatctic (structures in word sequences) or prosodic (accent and intonation patterns marked by internal and boundary features, e.g. 'hat patterns' versus sequences of pitch peaks), with the latter two either operating separately or in conjunction.

**Hypothesis 4 – reduction of the spatial and temporal extent of gesture opening**
Vocoid openings in opening-closing gestures show reduction to [ə] and to [ə] deletion if they occur in unstressed function words, or to deletion of [ə] if it immediately follows the stressed syllable in content words. This reduces the vertical extent of the opening movement of prosodically weak gestures right down to zero. The reduction has a particularly high frequency of occurrence in the structure oral closure + schwa + nasal, because it prepares the way for further motor economy with favourable perceptual conditions under Hypothesis 5.

**Hypothesis 5 – exclusion of the apical articulator in nasal and plosive production**

***(a) in apical nasal closing of word-final gestures under Hypothesis 4***
If [ə] under Hypothesis 4 occurs in a word-final gesture that opens from an oral occlusion and closes into a second oral occlusion made by the tongue tip for a nasal consonant, the gesture-initial articulator (lips, tongue dorsum) is maintained across the whole reduced gesture. This is not the case if the second closure opens into another vocoid in the same word.

If under Hypothesis 4, the nasal closing follows an opening from a fricative stricture, two separate stricture types have to be maintained in the gesture, and they stay associated more readily with two separate articulators, i.e. the frequency of place assimilation is expected to be lower after fricatives than after plosives and nasals.

***(b) in apical opening of word-final gestures under Hypothesis 4***
If the apical closure comes first – otherwise with the same conditions as in (a) – the gesture-final articulator establishes closure throughout the whole reduced gesture (e.g. *einem*, *jedem* get a single labial closure, but *regem* gets a sequence of dorsal and labial gestures).

***(c) in apical closing of word-final gestures before labial or dorsal opening of following word-initial gestures***
The word-initial gesture determines the articulator for gesture concatenation, either to establish a closure throughout between plosives and nasals, or for a closure-fricative stricture sequence.

The word-internal cohesion in (a) and (b) is much greater than the cohesion across words. So (a) and (b) are expected to be more frequent than (c). In (c), the syntactic as well as the prosodic grouping and the resulting degree of cohesion across the two gestures, according to Hypothesis 3, determine whether the apical articulation is carried out or not.

**Hypothesis 6 – no elimination of apical movements in fricatives**
Perceptual salience prevents the assimilation of apicals to labials and dorsals in fricatives, as against the processes in plosives and nasals under Hypothesis 5.

**Hypothesis 7 – velic control for plosives in nasal environment**

***(a) short-closure-duration lenis stops before nasals***

*(a1) in gestures under Hypothesis 4*
Velic lowering is advanced to such an extent that the stop closure is not only progressively shortened but eliminated altogether. If this change of velic control is added to articulator adjustment under Hypothesis 5 (a), (b), the degree of reduction in opening-closing gestures of the type plosive + schwa + nasal is increased beyond the effects derived from Hypotheses 4 and 5, presupposing raised values from the conditions of Hypotheses 1 to 3. Thus it is to be expected that the relative frequency of stop nasalization in the total sample of this gesture type is lower than the relative frequencies of schwa deletion and place assimilation on the same undifferentiated sample.

*(a2) in gesture concatenation within words*
The same nasalization process occurs as in (a1), e.g. in *Agnes*, but without articulator adjustment because of the gesture-initial position of the apical articulation.

***(b) long-closure-duration fortis stops before nasals***
Due to the long temporal safety zone in the fortis plosive occlusion the process can only be expected in unstressed function words with strongly reduced stop durations, under Hypothesis 4 and across word boundaries.

***(c) short-closure-duration lenis stops after nasals***

*(c1) in gesture concatenation within words*
Velic lowering is continued to such an extent that the short lenis stop closure is eliminated completely, e.g. in *Novem<u>b</u>er*. Since this process involves the perceptually more salient initial part of a gesture it is much less likely to occur than the reductions under (a).

*(c2) in gesture concatenation across words*
If the articulatory structures do not just span two gestures but also two words, the probability of occurrence is still further reduced. Conditions of Hypotheses 1 to 3 are essential, especially the occurrence of unstressed function words in first and/or second position of the gesture concatenation.

***(d) long-closure-duration fortis stops after nasals in gesture concatenation across words***
Due to the temporal safety zone in the fortis plosive occlusion (as under (b)), combined with perceptual salience (as under (c)), the process can only be expected in strongly cohesive structures of unstressed function words in a very relaxed and informal speaking style, if at all.

***(e) velic and glottal control for stop closure in bilateral nasal environment***

*(e1) inside words*
To produce a plosive in bilateral nasal context would necessitate a quick raising and lowering of the velum, whereas cut-

ting off the airstream at the glottis, or reducing it to very low-frequency pulsing, is much more economical, since the lowered velum position can be maintained, and it is equally effective in the signalling of an interruption to the listener. So glottalization can be expected in all the nasal + stop + nasal gesture sequences coming under Hypotheses 4 and 5 (a), (b) for both lenis and fortis plosives. However, lenis plosives can also come under the combined effects of Hypotheses 7 (a) and (c), resulting in a very long nasal, as in *Stunden*. This reduction will be rare for fortis plosives, and depends on the conditions of Hypotheses 1 to 3, especially on unstressed syllables and function words.

*(e2) across word boundaries*
The process also applies to this structure.

**Hypothesis 8 – reduction of the spatial and temporal extent of gesture closing**

***(a) lenition of obstruents between vocoid openings of successive gestures***
Due to stricture undershoot in gesture concatenation**,** plosives and fricatives may be realised as approximants, and there may also be a change of phonation from voiceless to voiced, because of a general weakening and shortening, particularly in unstressed syllables.

***(b) vowel nasalization and nasal consonant reduction***
A nasal consonant leads to a long articulatory component of nasalization across neighbouring sonorants, due to temporal constraints in the movements of the sluggish velum. In addition, the oral occlusion of the nasal consonant may be weakened through stricture undershoot, resulting in turn in distinctive vocoid nasalization. The status of this process is quite extreme in German and heavily depends on the conditions in Hypotheses 1 to 3.

***(c) geminate reduction***
When the onset of an oral opening and the offset of a preceding oral closing gesture have the same articulatory configuration we have the case of geminates. The timing of maximal constriction in this gesture concatenation may be changed in such a way that the offset of

the closing gesture becomes the onset of the opening gesture, thus reducing the temporal as well as spatial extent of gesture closing. This is a general tendency in German, but happens most frequently in consonant clusters, especially in unstressed syllable and function word sequences. Due to the cohesion condition of Hypothesis 3, geminate reduction is more frequent inside compounds than across separate content words.

**Hypothesis 9 – lateral reduction**
Lateral strictures are very special articulatory configurations in basically central opening-closing gestures and tend to be adjusted to the dominant central stricture dynamics in low articulatory energy environments, particularly in gesture-final position of unstressed syllables and function words.

**Hypothesis 10 – word-initial vowel related glottal features**
The word-initial boundary marker of vowel onset varies along an intensity scale from a strong glottal stop to weaker glottalization to wave amplitude modulation to complete disappearance. The degree of strength is a function of the conditions in Hypotheses 1 to 3, especially internal place in a prosodic phrase and enclitic function word junction lead to a weakening and elimination of glottal vowel onset.

## 2.3 Results

The details of the analysis of vowel and consonant related glottal stop and glottalization, of vowel modification and deletion, of place assimilation, lenition, nasalization, lateral and geminate reduction, and of the manifestation of function words may be found in Kohler (this volume, pp. 33ff), Rodgers (1999), Kohler and Rodgers (this volume, pp. 97ff), and Wesener (1999, 2001, and this volume, pp. 125ff). The following paragraphs summarize the general tendencies in relation to the hypotheses.

With the exception of geminate reduction according to hypothesis 8 (c), all phrase-level phonetic phenomena of the *Kiel Corpus of Read/Spontaneous Speech*, approached by Hypotheses 4 to 10, show a strong speaking style effect. Moreover, all data under Hypotheses 4 to 10 are influenced by the stress and function word conditions of Hypothesis 2, as well as by the cohesion factor of Hypothesis 3.

The progressive reductions of plosive + schwa + nasal consonant gestures according to Hypotheses 4, 5 (a), (b), 7 (a), (b), (e) are inversely related to their frequencies of occurrence. The different treatment of lenis and fortis plosives as to nasalization and glottalization has also been confirmed by their frequencies in the two data bases. The greater cohesion of the word-internal gestural structures under these hypotheses and the weaker cohesion across words in nasal/plosive + plosive/nasal under Hypotheses 5 (c) and 7 (c), (d) are mapped onto higher versus lower frequency distributions. The greater cohesion in word-internal, as against word-final or inter-word, gestures also manifests itself in more frequent schwa deletions in word-medial then in word-final syllables, and in the usual preservation of word-final schwa.

The absence of apical place assimilations in fricatives according to Hypothesis 6 has also been confirmed. Lenition of obstruents as well as nasal consonant, geminate, and lateral reduction show the regularities of Hypotheses 8 and 9, although the frequencies, except for geminate reduction, are quite low compared with those for Hypotheses 4, 5, 7, especially 4 and 5 (a), (b).

Overall schwa elision in plosive + schwa + nasal gestures reaches more than 90%, concomitant place assimilation is a little lower, but nasal assimilation of plosives is below 40%. This supports the view that the integration of plosive/nasal + schwa + nasal consonant gestures into extended oral closures, is being lexicalised in modern German, although historically it was a phrase-level phonetic reduction of an opening-closing gesture. The statistical evaluation of the data shows, however, that today the forms containing schwa and two articulators are elaborations, rather than starting points for reductions. The heuristics of the canonical base forms containing the more elaborated citation form feature, which was used in data processing and data bank searches along the 10 Hypotheses, made it possible to gain this insight into the relationship between lexicon and phrase-level phonetic reduction and elaboration of the extremely common plosive/nasal + schwa + nasal gestures in the German vocabulary.

# 3 Other investigations into phrase-level phonetics and future tasks

Over the past ten years or so an increasing interest in phrase-level phonetic phenomena, especially related to spontaneous speech, has developed world-wide. On the one hand, there has been the collection of large computer-accessible data bases, such as the Scottish Map Task Corpus, the Italian AVIP data base, the German VERBMOBIL corpus, the American ATIS and TIMIT corpora, and many more, which, in a number of cases, were triggered by speech technology applications. Centres for administering and agencies for distributing these data bases have been set up, such as the Bavarian Archive for Speech Signals in Munich, ELRA in Paris and the Linguistic Data Consortium in Philadelphia. The data recording standards and scenarios, the processing formats, the types of annotation (orthographic, phonetic, phonological, prosodic, grammatical), the transcription systems and conventions, the data labelling and analysis platforms, the data bank generation facilities, the search procedures vary enormously so that the application of the same phonetic questions in a multilingual perspective across several of these data bases is at least very difficult and full of problems for comparative data assessment.

The Kiel data acquisition and analysis platform, briefly presented in 2.1 as the basis for the investigations reported on here, constitutes a well integrated system from the recording of speech via its preprocessing, labelling, signal analysis, data bank construction to specific hypothesis-driven search operations on the data. It has been the goal of this speech processing environment not simply to provide a large structured data base as a scientific tool for the phonetic community, but to start using this tool immediately to find answers to questions of segmental and prosodic patterning in different speaking styles. The tool has been focussed on German, but it has also been applied to Russian, Italian and French, and the aim is to extend the analysis of phrase-level phonetics within this framework to as many and as diverse languages as possible.

Besides the preoccupation with large data collections in linguistics, there has also been a growing awareness of the need to move phonetic investigation from scripted laboratory speech to spontaneous speech. But the questions that researchers have had in mind first and foremost in this shift of focus have been centred on prosody, dysfluency and discourse interaction. Segmental aspects at the level above the phonological word and

questions of sound variability at the utterance level under the varying conditions of speech communication in different languages have roused much less interest in this new orientation towards spontaneous speech. And if phrase-level segmental questions were included they were frequently approached with an old methodology due to equating connected read speech with spontaneous speech. This was well illustrated by the international Symposium "Sound Patterns of Spontaneous Speech", organised under the auspices of the European Speech Communication Association by the Laboratoire Parole et Langage at Aix-en-Provence in 1998. The programme was dominated by the established paradigms of word phonology and phonetics, especially in the section on perception, which had no relation to spontaneous speech at all.

The two symposia organised by IPDS Kiel with German Research Council funding – "Sound Patterns of Connected Speech: Descriptions, Models and Explanation", 1996 and "Patterns of Speech Sounds in Unscripted Communication: Production, Perception, Phonology", 2000 – show that research into spontaneous speech sound patterns is gaining more wide-spread momentum and is being applied to an increasing number of languages, with a progression from the first to the second conference. Among papers of the second conference, published in the Journal of the International Phonetic Association, volume 31 (2001), the ones by Barry and Andreeva on Bulgarian, Russian, Czech, Polish, Italian and Greek, by Engstrand and Krull on Swedish, and Nikolaidis on Greek present spontaneous phrase-level phonetics within a theoretical framework that corresponds to the one adopted in the research at IPDS Kiel: these authors also admit categorical gestural reorganization, beyond gradual phasing relations, as well as more global units into their accounts of phrase-level phonetic reductions. That is, they, too, realise the need to overcome the narrow constraints of the model of Articulatory Phonology. Some of their data also testify to the same general reduction tendencies in these languages as have been found for German, especially as regards the simplification of opening-closing gestures in the extents of their opening and closing ('schwa-isation', lenition), the exclusion of the apical articulator and the occurrence of nasalization.

It is a task for the future to draw on the data resources of many more languages within a more unified processing environment and data bank structure in order to advance the theory, method and description of phrase-level phonetics of individual languages, of their dialects and regional as

well social accents, of language groups and of human language in general. By gaining greater insight into phrase-level variability under the various conditions of speech communication in a multilingual and multidialectal perspective, we will be able, in explanatory ways, to assess cross-linguistic similarities and differences of phrase-level phonetic patterns, to derive typologies and universal tendencies, and to connect synchronic phonetic variation with historical sound change.

# References

Barry, W. and B. Andreeva (2001). Cross–language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association 31*.

Browman, C. P. and L. Goldstein (1992). Articulatory phonology: an overview. *Phonetica 49*, 155–180.

Butcher, A. R., K. J. Kohler, and H. Kuenzel (Eds.) (1977). *Experimental Investigation into Coarticulation and Articulatory Control*, AIPUK 8.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology 13*, 139–169.

Byrd, D., A. Kaun, S. Narayanan, and E. Saltzman (2000). Phrasal signatures in articulation. In M. B. Broe and J. P. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V*, pp. 60–87. Cambridge University Press.

Engstrand, O. and D. Krull (2001). Simplification of phonotactic structures in unscripted Swedish. *Journal of the International Phonetic Association 31*.

Gnutzmann, C. (1979). Satzphonetik und Phonostilistik. In H. Vater (Ed.), *Phonologische Probleme des Deutschen*, Studien zur deutschen Grammatik 10, pp. 41–62. Tübingen: Gunter Narr.

Helgason, P. and K. J. Kohler (1996). Vowel deletion in the *Kiel Corpus of Spontaneous Speech*. In K. J. Kohler, C. Rehor, and A. P. Simpson (Eds.), *Sound Patterns in Spontaneous Speech*, AIPUK 30, pp. 115–157.

IPDS (1994). *The Kiel Corpus of Read Speech*, Volume I. Kiel: IPDS. Kiel CD-ROM #1.

IPDS (1995). *The Kiel Corpus of Spontaneous Speech*, Volume I. Kiel: IPDS. Kiel CD-ROM #2.

IPDS (1996). *The Kiel Corpus of Spontaneous Speech*, Volume II. Kiel: IPDS. Kiel CD-ROM #3.

IPDS (1997a). *The Kiel Corpus of Spontaneous Speech*, Volume III. Kiel: IPDS. Kiel CD-ROM #4.

IPDS (1997b). *xassp* User's Manual (Advanced Speech Signal Processor under the X Window System). In A. P. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), *The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results*, AIPUK 32, pp. 31–115.

Kohler, K. J. (1974). Koartikulation und Steuerung im Deutschen. In *Sprachsystem und Sprachgebrauch. Festschrift für Hugo Moser, Teil 1*, pp. 172–192. Düsseldorf: Pädagogischer Verlag Schwann.

Kohler, K. J. (1976). Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung. *Phonetica 33*, 1–30.

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: G. Narr.

Kohler, K. J. (1983). Prosodic boundary signals in German. *Phonetica 40*, 89–134.

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 Academic Publishers.

Kohler, K. J. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica 48*, 180–192.

Kohler, K. J. (1994). Complementary phonology: a theoretical frame for labelling an acoustic data base of dialogues. In *Proc. ICSLP94*, Volume 1, Yokohama, pp. 427–430.

Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Papers in Linguistics 11*, 21–33. Conference on 'The word as a phonetic unit', ZAS, Berlin, October 22–23, 1997.

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.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The Production of Speech*, pp. 217–245. New York, Heidelberg, Berlin: Springer.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling.*, pp. 403–439. Dordrecht: Kluwer Academic Publishers.

Nikolaidis, K. (2001). An elektropalatographic study of Greek spontaneous speech. *Journal of the International Phonetic Association 31*.

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.

Rodgers, J. E. J. (1999). Three influences on glottalization in read and spontaneous German speech. In K. J. Kohler (Ed.), *Phrase-level Phonetics and Phonology of German*, AIPUK 34, pp. 177–284.

Rodgers, J. E. J., P. Helgason, and K. J. Kohler (1997). Segment deletion in the *Kiel Corpus of Spontaneous Speech*. 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. 127–176.

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.

Wesener, T. (2001). Some non-sequential phenomena in German function words. *Journal of the International Phonetic Association 31*.