

The disappearance of words in connected speech

Klaus J. Kohler

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

1 Introduction

The category of the word is well established in meta-language pursuits, especially in linguistics. It is the basis for the development, over centuries, of the methods of lexicography, which have produced various types of lexica, including pronunciation dictionaries. The latter take the concept of the canonical phonetic representations of word citation forms in a language as their point of departure. At least some of them also list phonetic variants, which may occur in different utterance contexts (phonetic environments, levels of style). The pronunciation dictionaries presently available for various languages, differ enormously in the extent of taking the phonetic variability of words into account. At one end of the scale we find the DUDEN lexicon for German [1], which does not provide any contextual variants at all; at the other end are the English reference works by Jones/Gimson/Roach [9] and Wells [37], which give detailed information on phonetic variation of words in utterances. WDA (*Wörterbuch der deutschen Aussprache*) [39] is located in between, but closer to DUDEN.

This concept of word pronunciations and especially of canonical citation forms is built on the idea of the independent existence of individual words in an utterance, which in turn results in the assumption that words are units that can be defined and delimited phonetically. The word thus also constituted the frame of reference for the development of phonology: phonemes are the sound units that differentiate words, and boundary signals mark their beginnings and ends phonetically. As a consequence of this focus on word phonology the level of segmental sentence or utterance phonology was largely excluded from the study of sound patterns in languages.

Typical, often cited examples of phonetic markers for morphological structure are the palatal fricative in German "*Frauchen*" (noun "*Frau*" + diminutive "*-chen*") vs. the velar fricative in "*rauchen*" (stem "*rauch-*" + verbal ending "*-en*") and the dark lateral in (Southern British) English "*coolish*" (adjective "*cool*" + loose derivative "*-ish*") vs. the clear one in "*foolish*" (stem "*fool-*" + integrated suffix "*-ish*"). Lehiste's classic study of 'internal open juncture' of 1960 [25] also belongs to this field of word phonology (e.g. "*nitrate*" vs. "*night-rate*" or "*a name*" vs. "*an aim*").

Nevertheless it has been known from historical linguistics for a long time that potential phonetic boundary markers for word separation may be ignored at any time. In this connection we may compare "*Natter*" and "*Otter*" in German, and refer to English "*adder*", "*apron*" (as against "*napkin*"), or - with the opposite direction of sound change - "*newt*" (besides "*eft*"), "*nickname*" (as against "*eke*"): in all these cases the sequence of indefinite article and noun receives a new phonetic parsing. If modern linguistics had been dominated less by English and more by e.g. French the search for phonetic word indices would probably never have arisen. French, as a syllable-timed language with no lexical stress, lacks the phonetic marking of word units to a far greater extent than the stress-timed languages English and German, and consequently word puns abound. Here is a typical example:

*"De quelle couleur est toujours un coffre-fort quand on le vide?"
"Il est tout vert". - "Il est ouvert."*

This word orientation also determined a phonetic research paradigm which highlighted the word frame in experimental analysis still further, e.g. by the use of systematically varied nonsense words or of word contrasts in a constant utterance environment of the type “*It’s a (Say...) again.*” Although it is a reasonable assumption that the word is a language reality, at least for speakers, it reaches different degrees of awareness according to the demands of the communication situation, i.e. the word as a unit of speech will be particularly prominent in data collection under lab conditions, but far less so in spontaneous interchange. An investigation of the former kind uncovers coarticulatory effects and assimilations that stress the integrity of word units much more strongly, e.g. in the lack of complete labial/dorsal assimilation of coronal plosives and nasals at word boundaries, as in English “*hat pin*” or in German “*Schrottplatz*”.

The experiments by Nolan [27] and Kühnert [24] using EGP and EMA techniques, respectively, are cases in point. Having investigated apparent place assimilations Nolan went one step further in the interpretation of his data by proposing that differences in lexical phonological form always result in distinct articulatory gestures, even if overlapped and/or reduced or not discernible in the instrumental record. This is the complete reification of the phonetic word. But there is a good deal of evidence that the word boundary can be overridden in such cases, resulting in complete assimilation, especially frequent in, but not limited to, the reduction of function words, as in German “*mit dem*” with [mipm], [mimm] or [mim], besides [mitm].

The few examples quoted so far will have demonstrated that words may be identifiable as phonetic units but that they may also lose this phonetic identity, either by the change or the disappearance of boundary signals or by the entire fusion with other words. The phonetic manifestation of words thus oscillates on a scale from distinct separation to complete integration. The conditions for this phonetic variability of word identity depend on a number of factors:

- the general articulatory strategies in human language
- the individual language concerned
- the word class as well as the morphological and syntactic structures
- sentence accent, position in the utterance and general phonetic environment
- and, above all, the demands of the communicative situation as regards the balance between articulatory ease and auditory distinctivity, which is adjusted differently for different speaking styles - lab speech, read speech, spontaneous dialogue etc.

In order to be able to come to grips with this question of the word as a phonetic unit it is essential to go beyond the prevalent pattern of word phonology and move on to a consideration of the sound structures above the word at the sentence and utterance levels. This has been a focus of research at IPDS Kiel since the early 1970's and was mirrored in a German Research Council funded International Symposium on “Sound Patterns of Connected Speech”, organised at Kiel in June 1996 [36]. As regards German, there is now a sufficiently large, phonetically annotated acoustic data base of read and spontaneous speech, of altogether 70,000 running words, completely transcribed segmentally and in part also with prosodic labels: ‘The Kiel Corpus of Read/Spontaneous Speech’ on four CD-ROMs so far [4,5,6,7,21]. Together with a data bank environment and appropriate search as well as analysis tools it provides the necessary facilities [8,20,22,28] for large-scale corpus studies of connected speech processes in German [3,10,12,13,14,15,16,23,29,30,33,34]. A research grant from the German Research Council that has recently been allocated to IPDS for this type of investigation will allow us to exploit these speech resources more fully.

Moreover, this Conference at ZAS testifies to an ever growing awareness of the need for phonetic analysis of connected speech, and the organisers are to be congratulated for their initiative to run it by the side of a linguistics meeting on the **phonology** of the word. Therefore I am particularly grateful to them for giving me the opportunity to hold up the flag for phonetics by inviting me to speak to you here today. Since the obliteration of words in speech is more interesting than their preservation I have chosen the former as my subject.

2 The disappearance of words as delimitable units in speech production

2.1 Function words: from separation to integration

2.1.1 Disappearance of syntagmatic and paradigmatic phonetic word distinctions

The interference with phonetic word identity is particularly frequent in (sequences of) function words, e.g. in German

"Hast du einen Moment Zeit?" [haspm mom'en ts'ait]

"Hast du den Bericht über die letzte Sitzung endlich geschrieben?" [haspm bəʁ'ɪçʃt]

The same phonetic form [m] in the strongly reduced sequence of three function words *"hast du einen/den"* can be uniquely identified with *"einen"* in one context and with *"den"* in another, although the solely remaining nasal (with labial adjustment to the following consonant) can no longer trigger the phonetic identification of the word. The separation of these words is further hampered when instead of [pm] a glottalized nasal [m̥] is produced, which signals the article and the plosive residual of *"du"* at the same time.

But the reduction can go further and eliminate all traces of *"du"* in [hasm mom'en ts'ait], with a syllabic nasal, which may in turn follow the general German geminate reduction, especially in unstressed position and fast speaking rate, resulting in [has mom'en ts'ait], where the reflex of *"einen"* has also disappeared in the phonetic manifestation. The verbal paradigm as well as the idiomatic phrasing make the decoding of the intended meaning of the utterance unique, and the listener therefore does not depend on the signal detection of every word.

2.1.2 Emergence of new words through syntagmatic fusion

The disappearance of words in context is not restricted to the loss of all phonetic traces but may also take the form of the appearance of new lexical items through the complete fusion of others. This is particularly common for prepositions + articles, as in French *"au"*, *"du"* or German *"im"*, *"ins"*, *"zum"*, *"zur"*. In today's usage, German *"er geht zur Schule"* and *"er geht zu der Schule"*, *"er kommt zum Schluß"* und *"er kommt zu dem Schluß"* have different meanings although both forms are historically related on a scale of articulatory reduction.

Similarly, subject pronouns in enclitic position to function verbs form a scale from separation into two items to fusion into a single new one in e.g. German *"haben wir"*, *"sind wir"*, *"hat er"*, *"habt ihr"*:

[ha:bən vi:ə] [zɪnt vi:ə] [hat ʔeɐ̯] [hapt ʔi:ə]
 [ha(ɪ)m vɪə] [zɪm(p) vɪə] [hat (ʔ)ɐ̯] [hapt (ʔ)ɪə]
 [ha(ɪ)m vɐ] [zɪm vɐ] [hat^h ɐ] [hapt^h ɐ]
 [ham ɐ] [zɪm ɐ] [hat ɐ] [hapt ɐ]
 [hamə] [zɪmə] [hadə] [haptə].

The same subject pronouns in proclitic position and the indirect object *"ihr"* (in e.g. *"er hat ihr geholfen"*) reduce less, the possessive pronoun *"ihr"* (in e.g. *"sie hat ihr Kleid gewaschen"*) least:

in these cases fusion does not occur. So the disappearance of words in context and the appearance of new ones is not only situationally determined but also morphologically and syntactically.

2.1.3 Incomplete word fusion

A third type of the integration of words is illustrated by some of the reduced phonetic variants in “*die können wir uns abholen*” and “*die könnten wir uns abholen*” vs. “*die können uns abholen*” [dɪ kœmʏ ɱʏ ʊnʏs] and [dɪ kœmʏ ɱʏ ʊnʏs] vs. [dɪ kœnn ʊns].

Here words neither disappear without trace nor are they fused to new units: on the one hand, the sequential articulatory movements are greatly reduced, but, on the other hand, phonetic components of velarization, glottalization, nasalization etc. are kept as long residual traces of the eliminated elements overlaying the remaining ones. In these instances the tendency towards integration by articulatory fusion is counteracted by the opposite tendency to maintain the phonetic identity of the word through **articulatory prosodies** [11].

An extreme case of this is found in the series of four function words of “*nun wollen wir mal kucken*”, for which the Kiel Corpus of Spontaneous Speech provides [nū: ʊnʏ ẽ ɱa kʰʊkɳ] (OLVg122a009); see spectrogram in [36], p. 2.

The four initial function words of the sentence “*da hat er auch keine Zeit*” (with the sentence accent on “*Zeit*”), which are clearly separated in the precise pronunciation [da: hat ʔe:ʔ ʔaʊχ], may reach the stage of complete fusion in [da:de ʌχ], where [da:de] approaches a new lexical (phrasal) item as part of a paradigm [da:βɪç] [da:stə] [da:tsə] [da:mə] [da:ptə] [da:mzə]. But the componential element of breathy voice control may be kept as a residue of /h/, superimposed on the vowel of [da:] preserving the identity of the word “*haben*”, as in the following example from the Kiel Corpus of Spontaneous Speech: “*da haben Sie auch wieder recht natürlich*” (HAHg071a019); see spectrogram in [34], p. 140.

2.2 Componential residues of segmental deletions

In all these residue cases the componential features have to be represented in a phonetic transcription, even if it is basically segmental, because they mark phonological contrasts at the level above the word. In our labelling system in the Kiel Corpus, we have adopted the symbol -MA, inserted into the canonical transcription before symbolically deleted segments [21]. Its use may be illustrated by the following example (see spectrogram in [34], p. 157 and [11]).

TIS071a004 “*wahrscheinlich ein bißchen*”

canonical SAMPA v a:6 #S 'aI n l I C Q aI n+ b ' I s C @ n

variant SAMPA v a:6 #S 'aI n -MA l- I- C- Q- aI- n-m+ b -h ' I s C @- n

IPA [ʋaʃaɪn m bɪʃçɐn]

- The syllable l I C is characterized by palatality, i.e. by a high elevation of the tongue dorsum, which is obvious for I C but also applies to the clear (palatalized) l. I before C is, moreover, produced with a higher tongue position than before non-palatal consonants, e.g. in the suffix “-nis”. So the difference between I and C is one of vibrating and open glottis with very similar tongue heights; these phonation differences together with similar oral strictures generate laminal versus turbulent airflow at the tongue-palate opening, resulting in approximant and fricative articulation, respectively.
- l and I are articulatory opposites in their central and side tongue-palate contacts, which puts high demands on the execution of the speech gesture chaining.

- The tongue tip/blade gesture is subordinated to tongue dorsum and lip movements; therefore, under these sequential constraints, the palatalized **l** loses its central coronal contact in the dental/alveolar area by adjusting to the purely dorsal gesture of **l**. This is found generally in the suffix “-lich”, e.g. in “*selbstverständlich*”, “*natürlich*”, particularly when the words are unstressed and non-final in the utterance.
- In unstressed syllables all articulatory gestures are probably reduced in their magnitude, including subglottal pressure and glottal opening for **C**. The result is the transformation to an approximant with the possibility of voicing: **j**.
- The dorsality of the reduced final syllable may then also be extended to the preceding nasal, due to the higher rating of dorsum over tip/blade gestures in articulatory sequencing, resulting in a palatal.
- With the desynchronization of velic movement, especially between two nasals, i.e. before **m**, which originates from the reduction and assimilation of “*ein bißchen*”, the dorsal approximant is nasalized as well.
- If the closing of the lips for **m** occurs early enough there will not be an approximant stricture between the nasal of “*wahrscheinlich*” and the nasal **m**.
- So we end up with the pronunciation found in the spontaneous speech example as a consequence of natural constraints on articulatory gestures: a componential residue of palatality remains although segmental units corresponding to a canonical form can no longer be separated.

The application of **MA** is particularly important in the case of the deletion of a vowel as a voiced sonorant stretch at the segmental level. This will now be discussed with reference to variants of the word “*vielleicht*” in the Kiel Corpus of Spontaneous Speech. Among the high-vowel elisions, this word supplies a very high incidence of **MA** markings [3, 16]. So the analysis of the phonetic realisations of this item will be particularly informative from the point of view of spontaneous speech motor control. The following labellings (in SAMPA notation [38]) of the first, unstressed syllable will be considered: **f l l**, **f -MA l - l**, **f l - l** (see spectrograms in [3], pp. 122-127).

The gestural components that make up this speech unit are a labiodental stricture, a high front dorsal tongue position, a coronal closure with simultaneous lateral opening and a glottal abduction-adduction sequence. The precise temporal coordination of all these constituents is highly variable and continuous rather than discrete. In the hyper-version of the utterance the high dorsal position is maintained after the labiodental release, in turn followed by the tongue tip/blade and side gestures, and synchronised with voice onset. Deviations from this organization in the corpus data are:

- voice onset is delayed in relation to the labiodental release along a scale up to complete devoicing of the vowel;
- the onset of the coronal gesture is advanced in relation to the labiodental release along a scale up to complete disappearance of a separate vowel element; the syllable timing may otherwise remain unchanged (resulting in a syllabic lateral) or get shortened concomitantly (producing a non-syllabic lateral);
- the advanced timing of the coronal gesture may be combined with voice onset delay, again along a scale, resulting in more or less devoiced laterals;
- the high, front dorsal tongue movement may be kept in spite of the early coronal timing (resulting in palatalization within the **f l** cluster, particularly after an immediately preceding front tongue elevation, e.g. in the word “*nich(t)*”), or there may be early coarticulation with the diphthong **a l** onset of the subsequent syllable (and thus coalescence with the word-initial cluster **f l**).

All the realizations of “*vielleicht*” discussed so far are the result of temporal sliding between the coronal, dorsal and glottal gestures in relation to the labiodental release, and due to the continuous variation along these three timing scales there is a great variability in the recorded data. But there are also instances of this word in the data base that point to a different speech production strategy. It has to do with the articulations required for the sequence **I l** being opposites: high palatal dorsum elevation with front opening and side contacts for **I**, and with front closure and side openings for **l**. The articulatory transition puts high demands on speech motor control, especially under time constraints of fast speech and unstressed syllables, and there are three possible consequences:

- **I** is adjusted to **l**, which happens in the instances of early coronal gesture timing,
- there is a short period of all-round closure (corresponding to a segment **d**), for which there are also examples in the data base,
- **l** is adjusted to **I**: the coronal gesture is eliminated: **f I l- a I C t** [fr'eɪçt].

The latter process can no longer be subsumed under temporal sliding, but represents gestural reorganization: the tongue tip/blade movement is deleted from the articulatory plan, as in “*wahrscheinlich*”, discussed above.

Another few examples from spontaneous speech in the Kiel Corpus are to give illustrations of the variety of componential residues (see spectrograms in [22], pp. 14-17).

KAE g197a011 *könnten*
canonical SAMPA **k 9 n t @ n+**
variant SAMPA **k -h ' '9 ~ n- t-q @- n+**
IPA **[k'œ̃n̩n]**

- The first nasal consonant is deleted as a sequential element, but a residue of nasalization is still manifest in the preceding vowel as a componential feature.
- The plosive **t** is realized as glottalization somewhere in the sonorant context (vowel, nasal consonant), without a precise temporal and segmental alignment.
- In both cases the articulatory components require a non-linear symbolization, i.e. markers that do not receive durations:
 - ~ refers to nasalization
 - **t-q** to glottalization;
 - both are aligned to the same point in time as the following, non-deleted segment **n**,
 - indexing phonetic parameters in the segmentally labelled environment (further details in [21]).

HAH g074a010 *nicht zu spät*
canonical SAMPA **n I C t+ t s u: + S p 'E: t**
variant SAMPA **n I C t+ t s -MA u: -+ S p -h 'E: -'e: t**
IPA **[niç t̪^{wv}s^{wv}u̯^{wv}p'ert]**

- The voiced vocalic stretch of **u:** is absent;
- its lip rounding, and presumably its tongue position, remain as componential residues of labialization and velarization in the surrounding fricatives.

TIS g072a015 *kann Ihnen das*
canonical SAMPA **k a n+ Q i: n @ n+ d a s+**
variant SAMPA **k -h a n+ -MA Q- i:- n @ n-+ %d-n a s+**
IPA **[kʰan nʲ nas]**

- The segment **i:** is deleted;
- its dorso-palatal tongue elevation remains as a componential residue of palatalization in the nasal consonants.

HAH g074a000 *Universitätsstädte*
canonical SAMPA **Q U n I v E6 z I t 'E: t s #S t "E t @**
variant SAMPA **%Q U n I-i: v E6 z -MA I- t -h 'E: t s-S #%S t -h "E-"e: t-d @**
IPA **[ʔunivɛəzt'ɛ:tʃʰɛ:də]**

- The segment **I** after the fricative **z** is deleted;
- its dorso-palatal elevation remains as a componential residue of palatization in **z**, which, moreover, keeps its voicing as in intervocalic position although it now occurs before **t**.

Automatic labelling, such as the output of MAUS (see the contribution by Schiel and Kipp [35]), also ought to supplement the purely linear concept of the phonetic segment by the consideration of overlapping long articulatory/acoustic components, such as glottalization, nasalization etc.. Thus if MAUS labels a stretch of speech wave as **t @ m i: l I** (as in “*das paßt mir terminlich schlecht*”) it most likely does not capture the actual pronunciation adequately, because the nasal feature of the deleted nasal consonant **n** presumably lingers on in the nasalization of the vowel **i:** and the sonorant **l**, and the word-final fricative **C** leaves its trace in the word-initial **S**. In such a case the Kiel labelling would insert **~** in the first, **-MA** in the second instance. An automatic transcription has to do likewise, because it is only then that the symbolized pronunciation becomes empirically plausible; **t @ m i: l I** is not.

2.3 Content words: degrees of articulatory adjustment

The three types of interference with the phonetic unit of a word are not limited to function words. For example, in German numerals “-zehn” may be realized as **[tsn]**, and, over and above that, “-zehnhundert” (as in “*neunzehnhundert vierundneunzig*”) may even be pronounced **[tset]**, as long as the word refers to a year and “hundert” is not stressed. In the Kiel Corpus, for instance, we find the following variant (in SAMPA notation) for “(Mai) neunzehnhundert vierundneunzig” (**BACg142a005**); see spectrogram in [34], pp. 162.

n 'OY n t s e:- n- #h- "U- n- d- 6 t f'i:-'i:6 r- U- n t- #%n "OY n t s I C .

It is, on the one hand, a strongly reduced variant, linked to the citation form pronunciation

n'OYntse:n#h"Und6t f'i:rUnt#n"OYntsIC ,

on the other hand, it does not represent the end of the reduction scale because there may be further articulatory simplification, namely

- voiceless vowels in the voiceless obstruent environments
- **t** deletion before **s**
- deletion of nasal consonants and nasalization of the preceding vowels,

resulting in the variant

n 'OY ~ n- t- s e:- n- #h- "U- n- d- -MA 6- t f'i:-'i:6 r- U- n t- #%n "OY ~ n- t- s -MA I- C.

Filling in possible further variants between the canonical form, the corpus example and the most integrated pronunciation we get the following set of IPA-transcribed word sequences from most separated to most fused:

[n'ɔɪntse:nh,ʊndət fi:ʁʊntn,ɔɪntsɪç]
 [n'ɔɪntsənʰ,ʊndət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsŋʰ,ʊnnət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsŋ,ʊnnət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsŋənət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsnənət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsnnet fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntsnet fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪntset fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪnsət fi:ɛnn,ɔɪntsɪç]
 [n'ɔɪsət fi:ɛnn,ɔɪntsɪç]

The ordinal numbers ending in “-zehnten” provide further instances for the disappearance of phonetic words. Starting from canonical [tsentən], the following articulatory reductions occur:

[tsentn] with ə-elision,
 [tsenn̩] with additional glottalization instead of velic elevation to signal a stop articulation,
 [tsenn̩̤] with breathy phonation in the nasal instead, to mark this break,
 [tsenn] with the complete disappearance of the plosive reflex, which is possible in an unstressed syllable in nonfinal phrase position, e.g. before “November”, where we then get, for example, [dʁ'aitsen nov'embə].

This means that the cardinal and ordinal numerals in “dreizehn Novembertage” and “dreizehnten November” may coalesce.

In a labial context before, e.g., “Mal” we may find the variants with labial assimilation [tsem̩pm̩][tsem̩mm̩][tsem̩m̩m̩][tsem̩m̩].

“das hat er dreizehn Mal gemacht” and “das hat er zum dreizehnten Mal gemacht” may then coalesce in the form [tsem̩ maɪl]. The cardinal number can, however, have the further reductions [tse maɪl] and [tsm̩ maɪl], which seem to be impossible for the ordinal number. But the latter may be [tsm̩m̩ maɪl].

The disappearance of an independent phonetic word and the creation of a new lexical item is also illustrated by the greeting “n Abend” instead of “guten Abend”. This extreme reduction of an adjectival form is only possible in cases of semantic “bleaching”, as in this formula of phatic communion; it does not occur if the word retains its meaning, as in “guten Appetit”. A case from English would be “St. Paul” [sm̩] vs. “a saint man” [seɪnt].

Two examples of word fusion from the Kiel Corpus of Read Speech are (in SAMPA notation):

dlms 091: “geben Sie mir die Verbindung” g 'e: b- @- n- z i:- m i:6+

dlms 001: “morgen vormittag” m 'O6 ~- g- @- n- .

3 Balance between articulatory economy and auditory distinctivity as a function of the communicative situation

The examples presented in the preceding sections suggest that word production is a compromise between articulatory economy for the speaker and acoustic distinctivity for the listener. Economy of effort in speech production is governed by a number of anatomical, physiological and temporal constraints in the speech producing apparatus that introduce directionality into reductions, such that they are not chaotic. Not just any changes, but only certain types are possible, which occur over and over again in the languages of the world and in historical sound change. For instance the development of nasal vowels is tied to the position before nasal consonants, which are in turn deleted; stops may become fricatives and approximants, and the latter may even disappear in inter-sonorant position, but the reversal of this chain is not possible.

These physically constrained tendencies to reduce effort are in their turn controlled by linguistic structures at all levels, from phonology to syntax and semantics, and therefore have different manifestations and distributions in different languages, although basic types can be generalized. Furthermore the degree of articulatory effort is governed by the precision the listener needs in order to understand, and this need is different in different speaking environments, for acoustic reasons as well as for reasons of redundancy in form and content. This redundancy is determined by the common core of linguistic context and context of situation in the widest sense between speaker and hearer, ranging from world knowledge through culture and society to the individual discourse setting.

The balance between articulatory effort and perceptual distinctivity is thus solved differently in various communication situations (cf. Lindblom's H&H theory [26]). In the lab speech situation the effect of the principle of articulatory economy is small and consequently the preservation of word identity is much greater than in read texts and even greater than in spontaneous speech taking place within delimited scenarios. This means that the study of different speaking styles [12] may be expected to yield different frequencies and different degrees of articulatory reductions or reinforcements, and are consequently a research area of great potential for gaining insight into human communication, an area that has been too much neglected for too long to the detriment of linguistic science. Modern phonetics has the theoretical and methodological tools to get on with the task and to put spoken language performance into its proper perspective vis-à-vis the linguistic imperialism of written language competence.

Because of this tug-of-war between production effort and perception ease it is also an important and interesting question how listeners manage - or why they do not manage - to decode various forms of spoken language, which may, in the case of casual spontaneous dialogue, be extremely "distorted" from the point of view of canonical word forms. The examples quoted in this paper can all be understood immediately by native speakers of German in the contexts in which they are uttered; even the strongly reduced version of "*nun wollen wir mal kucken*", spoken by itself is quite intelligible. So listeners do not need complete phonetic signals for all the words that make up an utterance.

On the other hand, utterances that do contain all the phonetic word information may not be comprehensible because they lack the necessary (non-phonetic) context of situation cues. An example is the following German sentence (in IPA transcription without word divisions and with punctuation marks to indicate sentence prosodies):

[m'ɛ:əŋɐptəh'ɔɪ]?[n'ɛ:].[m'ɛ:kdəm,ɛ:ənh'ɔɪ],[ɛptəb'ɛ:tn].

German listeners are usually not able to decode it at all - or at least not without repetition - as the pronunciation corresponding to the spelling

"Mähen Äbte Heu? Nee. Mägede mähen Heu, Äbte beten."

The hearer thus gets along with a lot less phonetic word signalling, but also needs a lot more contextual cues; how much less of the one and how much more of the other in what phonetic, linguistic and situational contexts is a question to be answered by future research.

A further, very important factor for utterance intelligibility is its prosody. What may look like a list of unconnected words on paper (Chomsk's "*furiously sleep ideas green colorless*" phenomenon), may be a perfectly structured utterance when given the right temporal, accentual and intonational properties. Jokes and English crossword puzzles thrive on this. A German example is "*Theo der Kaffee*" [t'e:ɔdək'afe:], corresponding to the properly punctuated spellings "*Theo, der Kaffee!*" or "*Tee oder Kaffee?*", depending on timing, intonation and pausing; but under certain prosodic conditions the utterance may stay ambiguous. This phenomenon may also be exploited across languages, as in the following example:

<i>Un petit d'un petit</i>	[æptidæpti]	Humpty Dumpty
<i>S'étonne au hall</i>	[setɔnoɔ]	Sat on a wall,
<i>Un petit d'un petit</i>	[æptidæpti]	Humpty Dumpty
<i>Ah! degrés de folles</i>	[adəgʁədʃɔl]	Had a great fall,
<i>Un dol de qui ne sort cesse</i>	[ədɔldəkɪnsɔrsɛs]	And all the king's horses,
<i>Un dol de qui ne se mène</i>	[ədɔldəkɪnsmɛn]	And all the king's men,
<i>Qu'importe un petit d'un petit</i>	[kəpɔʁtæptidæpti]	Could not put Humpty Dumpty
<i>Tout Gai de Reguennes.</i>	[tugdæʒɛn]	Together again.

(Adapted from *Mots d'Heures: Gousses, Rames*", London: Angus & Robertson (1968))

The text on the left looks French, and it also sounds French with the segmental pronunciation, transcribed in the center column, and with the appropriate French utterance prosodies added to it, but it does not make sense in French, because it is simply a string of unstructured words. But anybody familiar with the English nursery rhyme in the third column will immediately recognise it as this little poem pronounced with a heavy French accent.

The following example provides a corresponding German version of an English nursery rhyme:

<i>Liter mies muffelt</i>	[l'i:temism'ʊfəlt]	Little Miss Muffet
<i>Satan atü fällt,</i>	[z'atʌnət'y:fəlt]	Sat on a tuffet
<i>Hie Dinge kurz und weh.</i>	[h'i:ɔŋək'ʊɛtsən'v'e:]	Eating her curds and whey;
<i>Sehr Kämme Piks beide</i>	[zɛrək'ɛmɔpɪksb'aidə]	There came a big spider,
<i>Ente satt Daunen bei Seide.</i>	[ʔɛntəz'atɔ'aʊnbaz'aidə]	And sat down beside her
<i>Unfrei den mies muffelt, oh weh!</i>	[ʔʊnfʁaidənmism'ʊfəlt ʔov'e:]	And frightened Miss Muffet away.

(Adapted from *Mörder Guss Reims*", London: Angus & Robertson (1981))

The foregoing discussion has made it quite clear that the word may but certainly need not be a phonetic unit. The word is very flexible in its phonetic manifestation, and it can therefore not be considered “the central phonetic unit”, as postulated by Tillmann’s paper at this Conference.

4 Conclusion and outlook

Our knowledge about words as phonetic units in lab speech is fairly comprehensive for quite a number of languages, including English and German in particular. Phonetics has of late also been able to come to grips with the scale of decreasing word signalling from read sentences to read texts and to different types of spontaneous speech, as the data presented and interpreted in this paper testify. They show that the realisation of words by speakers is a constant interaction between phonetic integration for economy of effort on the part of the speaker, and phonetic separation for distinctivity on the part of the listener.

But this domain of phonetics above the word still requires a great deal of research, and it needs, above all, a new paradigm [19] for asking questions about pronunciation in a language. **Word phonology** has outlived itself. We have to look much more closely at the regularities of production and perception processes at the **utterance level** in actual speech communication, and this goal goes beyond the word as a phonetic unit and beyond the collection of phonetic variants lexicæ, because we should not just deal with the question of how the words of a language are pronounced, we also need to give answers why the pronunciations are the way they are under the constraints of the utterance in communicative context. This scientific perspective also demands a thorough integration of the symbolic domain of phonological structures with the signal domain of phonetic speech dynamics. At IPDS Kiel we have been working very intensively on the question of utterance phonology and phonetics overlapping, and interfering with, word phonology and phonetics. Our German Research Council grant will allow us to continue this work within a framework of fundamental research to gain deeper scientific insight into how speech works. The focus is on German but we are ultimately aiming at a comparative treatment of European languages [2,17,18,31,32].

5 References¹

(AIPUK = Arbeitsberichte d. Instituts f. Phonetik u. digitale Sprachverarb. d. Univ. Kiel)

- [1] DUDEN: *Das Aussprachewörterbuch*. 3. Aufl. Mannheim/Wien/Zürich: Dudenverlag (1990)
- [2] Helgason, P.: Lenition in German and Icelandic. *AIPUK* 31, 219-226 (1996)
- [3] Helgason, P., Kohler, K. J.: Vowel deletion in the Kiel Corpus of Spontaneous Speech. *AIPUK* 30, 115-157 (1996)
- [4] IPDS: CD-ROM#1: *The Kiel Corpus of Read Speech, vol. I*. Kiel: IPDS (1994)
- [5] IPDS: CD-ROM#2: *The Kiel Corpus of Spontaneous Speech, vol. I*. Kiel: IPDS (1995)
- [6] IPDS: CD-ROM#3: *The Kiel Corpus of Spontaneous Speech, vol. II*. Kiel: IPDS (1996)

¹Graphic signal representations and speech output of utterances referred to in this paper can be found at the following URL: www.ipds.uni-kiel.de/examples.html.

- [7] IPDS: CD-ROM#4: *The Kiel Corpus of Spontaneous Speech, vol. III*. Kiel: IPDS (1997a)
- [8] IPDS: *xassp* (Advanced Speech Signal Processor under the X Window System) - User's Manual. Version 1.2.15. *AIPUK* 32, 31-115 (1997)
- [9] Jones, D.: *English Pronouncing Dictionary*. 15th ed. (P. Roach, J. Hartman, eds.). Cambridge: Cambridge University Press (1997)
- [10] Kohler, K. J.: Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51, 38-51 (1994)
- [11] Kohler, K.J.: Complementary phonology: a theoretical frame for labelling an acoustic data base of dialogues. *Proc. ICSLP94*, vol. 1, 427-430, Yokohama (1994)
- [12] Kohler, K.J.: Articulatory reduction in different speaking styles. *Proc. XIIIth ICPHS*, vol. 2, 12-19, Stockholm (1995)
- [13] Kohler, K.J.: The realization of plosives in nasal/lateral environments in spontaneous speech in German. *Proc. XIIIth ICPHS*, vol. 2, 210-213, Stockholm (1995)
- [14] Kohler, K.J.: Phonetic realization of German /ə/-syllables. *AIPUK* 30, 159-194 (1996)
- [15] Kohler, K.J.: Phonetic realization of /ə/-syllables in German. *AIPUK* 31, 11-14 (1996)
- [16] Kohler, K. J.: Articulatory reduction in German spontaneous speech. *Proc. 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from control strategies to acoustics*, 1-4, Autrans (1996)
- [17] Kohler, K. J.: Glottal stop and glottalization - A prosody in European languages. *AIPUK* 30, 207-216 (1996)
- [18] Kohler, K.J.: Glottalization across languages. *AIPUK* 31, 207-210 (1996)
- [19] Kohler, K.J.: Developing a research paradigm for sound patterns of connected speech in the languages of the world. *AIPUK* 31, 227-233 (1996)
- [20] Kohler, K. J.: Labelled data bank of spoken standard German - The Kiel Corpus of Spontaneous Speech. *Proc. ICSLP 96*, vol. 3, 1938-1941, Philadelphia (1996)
- [21] Kohler, K., Pätzold, M., Simpson, A. P.: *From scenario to segment - The controlled elicitation, transcription, segmentation and labelling of spontaneous speech*. *AIPUK* 29 (1995)
- [22] Kohler, K., Pätzold, M., Simpson, A. P.: From the acoustic data collection to a labelled speech data bank of spoken Standard German. *AIPUK* 32, 1-29 (1997)
- [23] Kohler, K.J., Rehor, C.: Glottalization across word and syllable boundaries. *AIPUK* 30, 195-206 (1996)

- [24] Kühnert, B.: Die alveolare-velare Assimilation bei Sprechern des Deutschen und Englischen: Kinematische und perzeptive Grundlagen. *Forschungsberichte des IPSK München* 34, 175-392 (1996)
- [25] Lehiste, I.: An acoustic-phonetic study of internal open juncture. *Phoentica* 5 (Suppl.), 1-54 (1960)
- [26] Lindblom, B.: Explaining phonetic variation: a sketch of the H & H theory. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*, 403-439. Dordrecht: Kluwer Academic Publishers (1990)
- [27] Nolan, F.: The descriptive role of segments: evidence from assimilation. In D. Ladd and G. Docherty (eds.), *Papers in Laboratory Phonology II. Gesture, Segment, Prosody*, 261-280, Cambridge: Cambridge University Press (1992)
- [28] Pätzold, M.: *KielDat* - Data bank utilities for the *Kiel Corpus*. *AIPUK* 32, 117-126 (1997)
- [29] Rehor, C.: Phonetische Realisierung von Funktionswörtern im Deutschen. *AIPUK* 30, 1-113 (1996)
- [30] Rehor, C., Pätzold, M.: The phonetic realization of function words in German spontaneous speech. *AIPUK* 31, 5-10 (1996)
- [31] Rodgers, J.: Vowel deletion/devoicing. *AIPUK* 31, 211-218 (1996)
- [32] Rodgers, J.: Vowel devoicing/deletion in English and German. *AIPUK* 32, 177-195 (1997)
- [33] Rodgers, J.: A comparison of vowel devoicing/deletion phenomena in English laboratory speech and German spontaneous speech. *AIPUK* 32, 197-214 (1997)
- [34] Rodgers, J., Helgason, P., Kohler, K. J.: Segment deletion in the Kiel Corpus of Spontaneous Speech. *AIPUK* 32, 127-176 (1997)
- [35] Schiel, F., Kipp, A.: Probabilistic analysis of pronunciation with MAUS. Paper at ZAS Conference on "The Word as a Phonetic Unit", Berlin Oct. 1997
- [36] Simpson, A. P., Pätzold, M. (eds.): *Sound Patterns of Connected Speech - Description, Models and Explanation*. *AIPUK* 31 (1996)
- [37] Wells, J. C.: *Pronunciation Dictionary*. London: Longman (1990)
- [38] Wells, J. C., Barry, W., Fourcin, A. J.: Transcription, labelling and reference. In: A. Fourcin, G. Harland, W. Barry, V. Hazan (eds.), *Speech Technology Assessment. Towards Standards and Methods for the EUROPEAN COMMUNITY*, 141-159 (1989)
- [39] WÖRTERBUCH DER DEUTSCHEN AUSSPRACHE (H. Krech et al., eds.), München: Max Hueber (1969)