ARTICULATORY REDUCTION IN DIFFERENT SPEAKING STYLES

K. J. Kohler
IPDS, Kiel, Germany

ABSTRACT

This contribution outlines six principles for the insightful analysis of articulatory reduction and discusses the results of a comparison of some connected speech phenomena retrieved from two labelled data bases of German: the Kiel Corpus of Read/Spontaneous Speech.

PRINCIPLES

The study of articulatory reduction should take the following principles into account:

- reduction of effort in connected speech
- listener orientation: auditory constraints
- gestural reorganization
- control of global functional coordinative structures rather than of individual anatomical articulators: cognitive constraints
- constraints imposed by the varying demands of different communicative situations: speaking styles
- phonetic, rather than phonemic, processing of large acoustic data bases within a framework of complementary phonology.

These principles will be discussed one by one in the above order with reference to some examples from English, but I will in the main draw on reduction data from German scripted and spontaneous speech that has been collected and phonetically labelled at IPDS Kiel. The scripted data consist of acoustic records of isolated sentences and two texts and are from 53 Standard German speakers of a North German variety [1]. The spontaneous speech corpus was recorded by another 26 speakers, with the same dialect background, in an appointment scheduling dialogue scenario of the VERBMOBIL project [2,3]. The signal and label files, supplemented by phonetic variants lexica of all the word forms occurring in the corpus, are available on two CDROM's [4,5]. As the labelled data are in a modified SAMPA notation [1], this too will be used for illustrations in this presentation.

REDUCTION OF EFFORT

Assimilations and elisions at word boundaries, as well as reductions in function words follow a principle of reduction of effort in connected speech. For instance tongue tip gestures are eliminated in favour of lip and tongue movements (e.g. *it's in that box* [pb], *not in that case* [kk]). I have argued in several publications, e.g. [6], that speech articulation is characterised by constantly ongoing lip and tongue dorsum movements, with tongue tip gestures riding on them at specific points, almost exclusively associated with consonants. The two types of movements are also linked with different muscle sets. Apical gestures in this view are the special, marked feature in speech production, which is, on the one hand, put to use in the functional domain of languages (inflexions in Indo-European languages, deictics, articles), and is, on the other hand, weakened to reduce articulatory effort, related to the coordination of tongue tip to simultaneously ongoing tongue body movements. The assimilation of apical stops and nasals (/t/, /d/, /n/) to labials and dorsals is thus referred to articulatory-physiological mechanisms, which at the same time exclude assimilations of labial/dorsal to apical and those between labial and dorsal.

In German, '/@/ + nasal' in word-final syllables is a particularly interesting example of this apical to labial/dorsal assimilation, going either from right to left or from left to right. In this context, /@/ may be deleted, i.e. the apical and labial/dorsal gestures may get closely linked in time and thus be affected by the articulatory reduction principle. So canonical ['a:b@nth] *Abend* may be realized as ['a:bmth] or ['a:th@m] *Atem* as ['a:pm]. For canonical sequences of 'lenis plosive + /@/ + nasal' the reduction may go even further, producing a nasal gesture throughout the cluster and even reducing it to short duration; so we may, for instance, get the set of pronunciations, from most elaborated to most reduced, ['a:bEnth], ['a:b@nth], ['a:bnth], ['a:bmth], ['a:mmth], ['a:mth] (the first one being a reinforcement from canonical /@/). All these forms occur in connected speech.

These synchronic processes are to be differentiated from diachronic ones in sound change (e.g. Lat. octo > Ital. otto), where the prime cause is auditory, not articulatory, in the transmission from one generation of speakers to another. This leads to the second principle.

LISTENER ORIENTATION

The speaker-related control in synchronic processes is checked by listener orientation.

(a) First of all the position in the syllable is relevant: the initial place requires greater distinctivity for the listener's discrimination than the final place, especially in stressed word-initial syllables. They are landmarks for the listener to decode the incoming signal and to relate it to word sequences at higher processing levels. This excludes the assimilation of word-initial apicals to preceding word-final labials or dorsals (English *hat pin* [pp] vs. *tip toe* [pt], German *Lippen* [pm] vs. *gib nicht* [pn]). The acoustic-auditory distinctivity of the release burst and the aspiration strengthen the initial position and preclude assimilatory reduction.

(b) But there are also listener-oriented constraints related to segment features, namely the manner features stop/nasal/fricative. Stops only assimilate if they are not released (English *apt* [pth], *act* [kth], *picked* [kth], but *he picked me* [k(p)m] *a good one*; German *Akt* [kth], but *Zwischenaktmusik* [k(p)m], *Beamter* [mth6], but *Beamten* [mpm]). This is so because only the unreleased plosive has little perceptual distinctivity, with rather small differences between [t], [p], [k]. And what is not well differentiated for a listener anyway can be levelled more easily in the speaker's attempt to save effort. Released and aspirated stops with distinctive local acoustic frication properties do not meet this requirement. Nasals, being far less clearly differentiated according to place of articulation than released stops, can be assimilated in word-final position (English *happen* [pm], *organ grinder* [gN]; German *geben* [bm], *legen* [gN]). Fricatives [f], [x] vs. [s], on the other hand, are acoustically as well as auditorily very distinct and are not assimilated (English *this form* [sf] vs. *that place* [pp], German *Schußfahrt* [sf] vs. *Rundfahrt* [mpf] or *Schrottplatz* [pp]).

GESTURAL REORGANIZATION

Gestures disappear, and there is gestural reorganization. This differs from Browman & Goldstein's position [7], according to which only the timing and the amplitude of gestures are changed in a very mechanical way on the articulatory surface, whereas the gestures themselves remain as such, and new gestures cannot be created in the articulatory execution of a gestural score. However, with the disappearance of the apical gesture in the change [mth6n] > [mpm] for German *Beamten*, the stop formation and release are also changed: they are exclusively effected by velic control. There is no way of subsuming this under the variables of timing and amplitude. This becomes even more evident in the change [nth@np] > [mpmp] of German *mit bunten Papier-*

schlangen. As early as the first nasal the articulators have to be instructed for a labial gesture, although the triggering element comes last. Simple passive adjustment through contiguity cannot explain this phenomenon; there has to be active reorganization. Moreover, instead of the assimilated inter-nasal [p] we also find either a glottal closure or glottalization interspersed into a continuous nasal, i.e. the labial closure and the velum lowering are maintained throughout the sequence [m...m]. This means that the phonatory break, signalling a stop to a hearer, is achieved through replacing a velic closure by a glottal one or by creak [8]. An articulatory action is thus transferred from one movable structure of the vocal tract to another for gestural economy [9], adhering to a principle of functional equivalence for the output to a hearer. Again, only gestural reorganization at a more central level can explain this phonetic process. From the only possible explanation of these empirical data follows the fourth principle.

FUNCTIONAL COORDINATIVE STRUCTURES AND COGNITIVE CONSTRAINTS

Degrees of reduction in accordance with the balance to be struck between reduction of articulatory effort on the part of the speaker and perceptual discriminability on the part of the listener are governed by a reduction coefficient, which controls whole sets of articulators forming a global functional coordinative structure. So, e.g., in the series of reductions from [mIthde:m] through [mItm], [mIpm], [mIbm] to [mIm] in the German phrase *mit dem Bus* [10] we are dealing with a progression through three successive domains: I. the reduction of opening-closing movements, II. the coarticulation between apical and labial gestures and the increasing reduction of the former, III. the progressive shortening of the oral and velic closure configuration. In spite of reorganizations within and between these three domains, they constitute a continuous scale of reduction, along which the reduction coefficient is located for a speech act, i.e. differently for different speaking styles, e.g. scripted vs. spontaneous.

This reduction coefficient is governed by higher-level cognitive processing as an essential prerequisite to speech production: the gestural adjustments are primarily not constrained by the vocal tract and passive changes in timing and amplitude, but are conditioned to a large extent by an adaptation, on the part of the speaker, to the acoustic-auditory needs of the listener in given communicative environments and to the semantic and syntactic demands of the utterance. So speakers reduce less in response to the request for repetition or in unfavourable contexts of situation, e.g. in noisy and more formal conditions, when failing to be understood is rated high (cf. also Lindblom's H & H theory [11]). Moreover, the semantic content of an utterance and its syntactic structure check the degree to which articulatory reduction can operate. Content words are not normally subjected to the same simplification as function words, unless they are weakened semantically at the same time, as in greetings and other sterotype formulae in phatic communion (compare English *kyou* as against *many thanks* and German *n'Abend* versus *guten Appetit*).

The reduction of function words is furthermore governed by word class (e.g. German *ihr* as a personal pronoun is reduced more than in the function of a possessive pronoun) and by position in a syntactic structure (e.g. enclitic *ihr* in German *habt ihr (den Film gesehen)?* is reduced more than proclitic *ihr* in *ihr habt (mich enttäuscht)*). In the German sentence *Er ist der, der der Sache am meisten schadet.* ("He is the one that endangers the cause most."), *der* refers to the demonstrative or the relative pronoun or to the definite article, and all three can be represented by the phonological form /d'e:r/, but the first is realised as [d'e:6], the second as [dE6], the third as [d6], with progressive reduction in accordance with decreasing prominence in different syntactic slots, for which the reduction coefficient is set at different points on the reduction scale (see also [12]).

SPEAKING STYLES

Articulatory reduction is more frequent and more extreme, the closer the speaking style is located to the informal and spontaneous ends of the formality and spontaneity scales. There are two ways in which spontaneous speech differs from reading style speech production with regard to articulatory reduction phenomena: either the degree of gestural levelling is increased to produce more extreme articulatory simplification, or the reduction features that are found in scripted speech have a higher frequency of occurrence, they turn up more readily. Both possiblities can be illustrated by a comparison of data from the two German corpora.

Spontaneous dialogue provides many examples that go beyond the rules set up for German on the basis of scripted speech [6]. The following two instances from the Kiel VERBMOBIL corpus [5,15], which can be multiplied manifold, illustrate the phonetic processes at work.

(a) The phrase *wahrscheinlich ein bißchen* ("probably a little") has the canonical citation form (in SAMPA trancription):

va:6#S'aInlIC QaIn+ b'IsC@n

but is realized (labelled) as:

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

where a symbol followed by '- space' means deletion, a symbol preceded by 'space -' insertion and 'symbol - symbol' replacement. So in this case the segments of the final syllable of *wahrscheinlich* are all deleted, with reference to the citation form, but their long component 'palatality' is preserved as a feature in the preceding nasal, which has palatal place of articulation. To mark this componential (prosodic) feature in a linear segmental transcription, -MA is inserted as a general marker that can be substantiated by subsequent speech signal analysis. -MA symbolizes componential features that are still present in spite of the disappearance of delimitable segments (e.g. nasalization, labiodentalization, vowel quality and duration residue, palatalization, velarization etc.). It receives its componential meaning from the segments marked as deleted in relation to the segment preserved before it. This subcategorization will eventually be performed automatically by computer programme based on distinctive feature representations of the segments.

(b) In the phrase *ich kann Ihnen das ja mal [sagen]* ("I can perhaps suggest this to you") the canonical form is:

QIC+ kan+ Qi:n@n+ das+ ja:+ ma:l+,

but its realization is labelled as:

Q- I C+ k -h a n+ -MA Q- i:- n @- n-+ d-n a s+ j a:+ m a: l+.

Here the high dorsum articulation for the first vowel in *Ihnen* is transferred to the alveolar nasal consonant following it. The apical nasal consonant articulation is maintained from the end of *kann* through *Ihnen* to the beginning of *da*, but inside it palatalization reflects the vowel of the function word *Ihnen*, although it is no longer realized as such.

These extreme reductions are either absent or very rare in speech reproduced from writing. There was only one speaker in the Kiel Corpus of Read Speech who provided instances of this type. For example his rendering of *morgen früh*, canonically represented as [m'O6g@n fr'y:], was [m'O~ fr'y:], with a nasalized vowel. So this speaker went beyond the well-known process of consonantal nasalization, found in scripted speech and derivable in the sequence [g@n]>[gn]>[gN]>[NN]>[N], by not only advancing the lowering of the velum in time but by changing the dorsum closing gesture as well. This informant is characterized by a general articulatory imprecision, which, in everyday interchanges, also prompts the re-

Table 1. Absolute frequencies of /@/ elision and various assimilation processes in '/@/ + nasal' of word-final syllables

(a) Read Speech		(b) Spontaneous Speech	
total number of words in corpus	*31,374*	*total number of words in corpus*	*9,291*
total number of sample words, incl. 52 compound double entries	*5,117*	*total number of sample words, incl. 16 compound double entries*	*1,605*
/@/ reinforcement	10	/@/ reinforcement	1
/@/ preservation	950	/@/ preservation	109
/@/ deletion	4,157	/@/ deletion	1,495
- place assimilation	1,146	- place assimilation	339
-- across word boundaries	52	-- across word boundaries	50
-- lenis stop nasalization	13	-- lenis stop nasalization	18
-- no lenis stop nasalization	824	-- no lenis stop nasalization	151
- no place assimilation	2,991	- no place assimilation	1,134
-- contextually not possible	2,414	-- contextually not possible	846
-- contextually possible, but not made	312	-- contextually possible, but not made	72
-- lenis stop nasalization	151	-- lenis stop nasalization	99
-- no lenis stop nasalization	333	-- no lenis stop nasalization	41
- treatment of /n/		- treatment of /n/	
-- deletion after nasalized stop	145	-- deletion after nasalized stop	93
-- deletion after canonical nasal	120	-- deletion after canonical nasal	119
-- preservation after any nasal, with place assimilation	314	-- preservation after any nasal, with place assimilation	136
- extreme reduction	20	- extreme reduction	22

quest for repetition more often than usual.

Although the two labelled corpora for German are of different sizes (31,374 word forms of scripted and 9,291 word forms of spontaneous speech) and from a different number (but homogeneous dialect group) of speakers, comparative relational statements about the occurrence of reduction phenomena are possible. I have selected the treatment of '/@/ + nasal' in word-final syllables, already mentioned. Table 1 provides an overview of the data, including non-final word components in compounds.

The proportion of the sample size to the total corpus size is comparable for the two corpora. Preservations and reinforcements of the vowel make up 7% of the sample in spontaneous and 19% in read speech, which is an indication that /@/ is dropped more readily in spontaneous interactions. Contrariwise, place assimilations of /n/ as a proportion of the total of possible place assimilation environments are 82% for spontaneous and 79% for read speech. It may thus be concluded that the two speaking styles do not differ in apical gesture reduction. However, if this parameter is looked at in combination with the nasalization feature in lenis stops there are clear differences: (a) nasalization is added to place assimilation in 11% of the possible cases of unscripted, but in only 2% of scripted speech; (b) in the absence of place assimilation, the incidence of nasalization is even more strikingly in favour of spontaneous speech: 71% vs 31%; (c) overall percentages of nasalization of lenis plosives are 38% vs. 12%.

The treatment of /n/ in 'nasal + /@n/' syllables is also characteristically different between the two speaking styles: in spontaneous speech the nasal is preserved - with place assimilation to the preceding nasal of any origin (canonical nasal or nasalized lenis stop) - in 39% of the cases, as against 54% in read speech. This reflects the greater tendency in spontaneous speech to reduce long syllabic consonants to short non-syllabic ones. These data also indicate that the reduction of 'lenis stop + /@n/' syllables is more likely to go beyond the place assimilation and nasalization stages in that speaking style. Finally, the proportion of more extreme reductions, beyond these well-established categories of /@/ elision, place assimilation, nasalization and syllabic nasal deletion, is greater in spontaneous speech; in reading style the relevant instances are largely due to the one speaker already mentioned: the example from his speech discussed above constitutes an extension of the reduction scale in 'lenis + /@n/' syllables at issue here.

The data suggest that articulatory features show different degrees of susceptibility to reduction. The elimination of a separate opening-closing movement in /@n/ syllables, i.e. a simple change of gestural timing, seems to be common in both speaking styles, but reading clearly restrains this tendency. Similarly, place assimilations reflecting gestural reorganization, although not as frequent as /@/ elision, are also quite regular when the special contextual conditions prevail, but the two speaking styles do not seem to differ. On the other hand, nasalization of lenis stops and reducing syllabic consonants, again due to gesture timing, are less common, and there is a large difference between the two speaking styles in the frequency of occurrence. The further reduction involving another gestural reorganization to eliminate a closing movement of, e.g., the tongue dorsum is even less likely to occur and closely linked to a spontaneous speaking style, even if it is introduced into reading by speakers who obviously have a less well defined separation of phonetic registers.

PHONETIC PROCESSING

The corpora of data used as a basis for the aforegoing discussion are labelled within a linear segmental phonemic framework. A powerful grapheme-to-phoneme conversion module within the RULSYS/INFOVOX TTS system [1,13] is used to automatically generate phonemic notations from the orthographic input (scripted text files read by the subjects and transliteration files of spontaneous recordings [14], respectively). After manual correction (appr. 3% error rate for running text), canonical transcription files result, which, together with the corresponding speech signal files are input to an adaptation of the KTH/ Stockholm MIX programme [13] for linear segmentation and labelling, with guidance from graphic signal plots (oscillogram, spectrogram) and acoustic output. The result is a canonical transcription file representing the actual pronunciation within a segmental description. All the labels of the canonical transcription are taken over, modified as deletions or replacements, if necessary, and supplemented by marking insertions. It is at this stage that additional phonetic labels are introduced at the subphonemic level (plosive releases, nasalization, creaky voice etc.). Many, but by no means all, of these markers are again segmental and linear (but cf. the use of -MA in the section on 'Speaking Styles').

This linear segmental frame allows the systematic and economical representation of lexical items in canonical citation form, to which actually spoken word forms can be related. This makes it possible to search label files generated in this way for such phonetic processes as assimilations and elisions, as exemplified in the preceding paragraphs. But in spite of these great advantages, which the linear approach offers, it also has serious drawbacks:

- Segments may not always be discernible in the signal as sequential elements, but their reflexes may nevertheless still be present as componential modifications of remaining segment strings, referable to such processes as palatalization, velarization, nasalization etc. The examples discussed in the preceding section illustrate this.

- If only the segmental deletions are marked there is a loss of contrastive phonological information because the signal contains more relevant features than this kind of symbolization represents, but the strictly linear segmental approach with a phonemic orientation is not capable of capturing this distinctivity.

As examples of this non-segmental residue of deleted segments are particularly common in spontaneous speech, a theoretical solution has to be found to deal with them adequately in the labelling of dialogue data bases. But the solution cannot consist of abandoning the linear segmental approach altogether and adopting a non-linear componential one instead because this theoretical reversal would forego all the clear advantages the linear concept has. So instead of an 'either-or' an 'as-well-as' is needed: besides individual segmental building blocks in their own right there are reciprocal influences on their concatenation, which manifest themselves in long phonetic components, even if the segments have been deleted in fast and reduced speech.

This integration of non-linear concepts into a linear phonological frame results in complementary phonology [15], a theoretical approach that requires both linearity and non-linearity - segments and components - for an adequate phonological representation of speech. The two concepts are only different aspects of the same phonetic-linguistic phenomenon. The segmental part establishes the relationship between phonetic manifestations in connected speech and canonical phonemic representations in a lexicon; the componential part takes into account the phonetic processes that characterize the concatenation of segments, especially in spontaneous speech and in its more numerous and more extreme articulatory reductions. This frame of complementary phonology provides a more adequate link between the symbolic and signal levels of speech analysis, interfacing symbolic descriptions of speech phenomena with events at the level of acoustic signals, which must in turn be related to the underlying articulatory processes.

For a more thorough pursuit of questions of segmental reduction in different speaking styles much larger data bases are required for more sophisticated statistical evaluation than have been analysed so far. And they have to be processed in such a way that they can give detailed phonetic information at the symbolic level in close correspondence with the acoustic signal and its various analyses. The symbolic representation should be within the framework of complementary phonology and the data entered into a data bank that allows the efficient and quick retrieval of acoustic signal data, also for further signal processing, in relation to symbolic strings contained in the label files and derived variants lexica. The Kiel data bases are structured along these lines and being put to use for the study of articulatory reduction in different speaking styles. What we need are data banks of this type for many languages in order to put the analysis of what has been termed 'phrase-level phonology' on a broader and comparative as well as typological level.

Although we can gain a great deal of insight into gestural dynamics under different speaking style conditions from the acoustic record, there are clear limitations as to what it can tell us. It is therefore mandatory to supplement our labelled acoustic data bases by labelled articulatory ones, but before this can be done successfully a methodology for representative data collection and processing will have to be developed.

The important conclusion to be drawn from the study of speaking style phenomena is the realization that the phonemic switches contained in a series of reduced forms along the reduction scale from least to most do not capture the essentials of what goes on in articulatory modification. It is the phonetic perspective associated with the biological and social constraints in communicative systems of sound producing humans that provides the answer to the question as to why articulatory reduction works the way it does.

ACKNOWLEDGEMENT

Part of the work reported here was carried out with financial support from the German Ministry of Education, Science, Research and Technology (BMBF) under VERBMOBIL contract 01IV101M7. My special thanks go to KTH/Stockholm for making MIX and related software available.

REFERENCES

[1] Kohler, K.J., *Lexica of the Kiel PHONDAT Corpus. Read Speech*, vols I, II, AIPUK 27, 28, Kiel: IPDS.

[2] Pätzold, M., Simpson, A. (1994), *Das Kieler Szenario zur Terminabsprache*. VERBMOBIL Memo Nr. 58, Kiel: IPDS.

[3] Kohler, K.J., Pätzold, M., Simpson, A. (1994), *Handbuch zur Segmentation und Etikettierung von Spontansprache - 2.3*. VERBMOBIL Technisches Dokument Nr. 16, Kiel: IPDS.

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

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

[6] 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/ Boston/London: Kluwer Academic Publishers.

[7] Browman, C.P. Goldstein, L. (1992), "Articulatory phonology: an overview", *Phonetica*, vol. 49, pp. 155-180.

[8] Kohler, K.J. (1994), "Glottal stops and glottalization in German. Data and theory of connected speech processes." *Phonetica*, vol. 51, pp. 38-51.

[9] Kröger, B.J. (1993), "A gestural production model and its application to reduction in German", *Phonetica*, vol. 50, pp. 213-233.

[10] Kohler, K.J. (1991), "The phonetics/phonology issue in the study of articulatory reduction", *Phonetica*, vol. 48, pp. 180-192.

[11] 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/ Boston/London: Kluwer Academic Publishers.

[12] Kohler, K.J. (1979), "Kommunikative Aspekte satzphonetischer Prozesse im Deutschen", in H. Vater (ed.), *Phonologische Probleme des Deutschen, Studien zur deutschen Grammatik 10*, pp. 13-39, Tübingen: Gunter Narr.

[13] Carlson, R., Granström, B., Hunnicut, S. (1990), "Multilingual text-to-speech development and applications", in W.A. Ainsworth (ed), *Advances in Speech, Hearing, and Language Processing*, pp. 269-296, London: JAI Press.

[14] Kohler, K.J., Lex, G., Pätzold, M., Scheffers, M., Simpson, A., Thon, W. (1994), *Handbuch zur Datenaufnahme und Transliteration in TP 14 von VERBMOBIL - 3.0* VERBMOBIL Technisches Dokument Nr. 11, Kiel: IPDS.

[15] Kohler, K.J. (1994), "Complementary phonology: a theoretical frame for labelling an acoustic data base of dialogues" *Proc. ICSLP94*, vol. 1, pp. 427-430.