# THE ORGANIZATION OF SPEECH PRODUCTION
## CLUES FROM THE STUDY OF REDUCTION PROCESSES

K. J. Kohler

Institut für Phonetik und digitale Sprachverarbeitung
Kiel, Germany

## ABSTRACT
Starting from various different views on levels of phonetic representation a unitary approach is presented which centres on the explanatory power of articulatory reduction processes and their auditory control.

## 1. LEVELS OF REPRESENTATION
The Journal of Phonetics devoted its July 1990 issue to the theme "Phonetic Representation" [1] and thus provided a good opportunity for the phonetics world to find out more about the heterogeneity of its views concerning the relationship between phonetics and phonology. The most encouraging outcome of this discussion is the unanimous realisation that phonology and phonetics need each other. This means that there has been progress in superseding not only the a-linguistic treatment of speech as purely physiological or acoustic events by the first experimental phoneticians, but also the exclusive reliance on symbolic phonological representations in the wake of Trubetzkoy's humanities/science dichotomy. What remains an open issue is the specification of the relationship. Is it to be seen as a mapping from one level of representation to another (following Keating [1, 321ff] or Pierrehumbert [1, 375ff]) or as an integration (following Browman and Goldstein [1, 299ff] or Fowler [1, 425ff])? How many levels of representation should there be: just two - phonology and phonetics, or four -

phonology, categorical phonetic representation, articulatory and acoustic parametric representations (Keating), or multiple, with as many phonetic representations as there are different measures of interest for a particular question in hand (Pierrehumbert)?

Intimately linked with the number of levels of phonetic representation is the question as to the ontological status of these levels. Phoneticians and phonologists are interested in finding out how speech communication works and therefore aim at the categories that make it possible, i.e. which are relevant for the language user and form the basis for communication. So they all tacitly or explicitly assume that the levels of phonetic representation are not just heuristic devices for making descriptions of data, but have a reality outside linguistics and phonetics in language and speech behaviour. It is at least doubtful whether this goal has been reached in all cases. Pierrehumbert's multiple phonetic levels are clear instances of descriptive frames, rather than inherent components in speech communication. Keating's categorical phonetic interface between phonology and parametric representations is likewise a surface descriptive device which does not address the question of phonetic explanation in phonology, e.g. with regard to the voicing contrast (cf. [2]).

But even Browman and Goldstein [1], in spite of their laudable attempt at integration in an articulatory phonology, do not go far enough, as Mattingly [1] so succinctly points out: "What is called for is a functional rather than an anatomical organization, one in which gestures are grouped according to the nature of their tasks." (p. 450) In such a functional approach, output considerations do actively constrain the processes of variation in speech production; variation with respect to speaking style and prosodic context does not simply follow from very general principles of gestural overlap and magnitude that are blind to their acoustic consequences [1, 303-4]. This becomes clear the minute we transgress the limitations of lab speech word contrasts and look at processes in connected, continuous and spontaneous speech, where gestures not only overlap to the extent that one becomes completely hidden by others, thus losing its acoustic consequences, but where gestures are constantly eliminated altogether, by comparison with more careful speech.

The most obvious case of heuristics is Ladefoged's discussion of the IPA classification and transcription framework [1, 335-346]. L. is quite right in stressing the pragmatic rather than theoretical basis of the IPA alphabet, but he is wrong in the conclusions he draws from this state of affairs for an anthropophonic approach. His argument runs as follows:
(1) Given the phonological basis of the IPA - whatever its precise nature may be - it is necessary to classify those sounds as different that are used in any one language to distinguish between words as well as those that are contextually conditioned allophones.
(2) When sounds occurring in different languages have to be classified the problem may arise of deciding whether they are the same, as, e.g., dental or interdental fricatives in different varieties of English.
(3) There are no language-independent criteria such as articulatory ease and auditory phonetic similarity that could solve this problem, because due to observer bias there is no principled way of setting up an auditory theshold or of measuring degree of articulatory effort in a general language perspective. What general reasons could we give for treating bilabial and labiodental fricatives as different, but dental and interdental ones as the same, both pairs involving small differences in place of articulation, but only the first being attested as a phonological opposition in languages, e.g. Ewe.
(4) It follows from (3) that all we can do is listing the speech sounds observed in all the world's languages.

In response, I would like to raise the following counter-points:
(a) Although establishing the segmental contrasts for word differentiation is an important phonetic task, it has to be supplemented by research in two further areas to be of validity for the elucidation of the speech communication process:
1. the analysis of contrasting segments in all possible phonetic parameters and in their differentiated influences on the segmental environment;
2. the manifestation of phonological oppositions in connected, continuous and spontaneous speech.
(b) It follows from (a)1. that, with regard to the fricative distinctions under (3), the comparably slight differences of place of articulation in the two pairs can have very different consequences on the inherent acoustic parameters of fricative intensity and spectrum as well as their temporal courses, with auditory results of different magnitudes (cf. [5], plate 11).

This point is further strengthened by the reference in [8, 79] that [ɸ] may alternate with [p] in Ewe, which means that the fricative can be related to a weakening of a closure in an articulatory reduction process, resulting in a weak fricative. This is similar to the change of intervocalic [b] to a voiced bilabial (frictionless) approximant [β] as a stylistic variant in casual German speech of, e.g., 'liebe' (*dear*), compared with a labiodental fricative [v] in 'primitive'. The very genesis of [β] under the conditions of reduction in less formal spontaneous speech of (a)2. rules out the development of a fricative.

(c) Thus the study of synchronic reduction processes in connected, continuous and spontaneous speech of a variety of languages will give us insights into general language-independent scales of articulatory effort, particularly when we can find independent physiological motivations. These synchronic data can be supplemented with data from diachronic sound change adding to their explanatory power.

(d) Phonemic oppositions do not all show the same stability. Some exhibit lower resistance to coalescence than others, for articulatory and/or acoustic-auditory as well as environmental (e.g. noise) reasons. Even when [ɸ] and [f] constitute separate phonemes for the differentiation of isolated words their auditory distance is smaller than the one that separates either of them from [s]. Similarly, [θ] and [f] are auditorily closer than [s] and either of these sounds, as can be judged from the sound change [θ] --> [f] in some varieties of English. It is indeed possible to make language-independent statements about the auditory distances or similarities of sounds, provided the techniques of investigation go beyond contrasting isolated words by linguistically biassed observers and include the native speaker reaction in articulation score type experiments [7], in direct similarity assessment, in slips of the ear during spontaneous speech, in the study of auditory constraints on articulatory reduction, for a variety of languages.

## 2. ARTICULATORY REDUCTION

I will now give a brief summary of reduction processes in German (for further details see [3, 4]) and draw conclusions from them for the organization of speech production and levels of representation. In the sequence of the preposition 'mit' (*with, by*) and the definite article 'dem' (dat.) in 'mit dem Auto' (*by car*) we can get the following series of reductions from most to least careful.

I.  mɪtʰdeːm          III. mɪpm
    mɪtʰdem                mɪb̥m

    mɪtʰdəm               mɪbm
    mɪtdm                 mɪmm
                          mɪm

II. mɪt⁻pːm
    mɪpːm

This series of segmental changes contains a great number of phonemic switches , and, therefore, the question arises as to whether we are here dealing with a single continuous scale of reduction and one large range of signal variability, related, say, to the most elaborate form (taken as the underlying abstract invariant) or whether this scale is discontinuous, with each phonemic change constituting a new invariant reference point for signal variability. The series can be divided into the indicated three groups according to speech production criteria. In I., we get a progressive shortening of the opening-closing movements between two oral closures, until there is no longer an open phase between them. In II., coarticulation occurs between two successive closure gestures (apical and labial/dorsal), and the apical one can be more and more reduced until

104

it is finally eliminated. In III., we then find a progressive shortening of the oral plus velic closure configuration, i.e., an increasingly earlier descent of the velum for the nasal release. With this shortening, passive voicing can continue through the stop closure phase, and eventually the complete blockage of air is eliminated altogether, the velum being lowered as early as the onset of the labial closure. Seen in this way, the phonemic switches do not throw any light on the significant changes in speech parameters, because they cut across them.

We thus have a continuous natural progression in the reduction of articulatory movements from the full form [mɪtʰ de:m] to the fully reduced one [mɪm]. Although there are a number of discrete phonemic changes along this articulatory continuum they do not capture the essentials of the speech production processes. In these cases, speakers do not set out to alter an underlying phonemic chain in a series of steps before they hand it over to the articulators for execution, nor do they select different forms from a lexicon, and the changes are not caused by simple peripheral vocal tract constraints either. It may be assumed that speakers start from a segmental representation of the full forms [mɪtʰ] and [de:m] taken from a lexicon, and attribute to the sequence a reduction coefficient at a high processing level before actual execution. The strength of this reduction coefficient determines the extent of articulatory smoothing, always beginning with the group I processes and moving further and further down the continuum with increasing strength. The phonemic representation of these reduction stages makes it possible to provide a first description of the phenomena, but it is misleading from a speech production point of

view, because it is also far too complicated. Instead, we can think of a general programme for articulatory reduction, whose components are specified and hierarchically ordered for a particular language or dialect (some may even be universal) according to the degree of reduction to be achieved. The components of this programme are triggered by the reduction coefficient. It is not necessary to specify all the possible types of assimilations etc. for individual segmental sequences, but quite general instructions along the lines given in the characterisation of the three groups of processes suffice in this general reduction programme, whose specific output depends on the application of the general phonetic rules to specific segmental sequences according to the strength of the reduction coefficient. Browman and Goldstein's articulatory phonology complemented by Mattingly's functional approach and extended to connected speech would provide a good basis for such a reduction programme.

But it would also have to be supplied with an auditory control component, because speakers not only control reduction with regard to the physiological and articulatory potentials contained in the dynamics of sound production but also take listeners into account and adapt to their needs [6] in two ways:
(a) Reduction processes are favoured that show a low degree of perceptual salience. That is the reason why apical fricatives, released apical plosives and syllable or word initial apical nasals and plosives are not assimilated in German (e.g. 'Beamte' [bə 'ʔamtʰ ə] vs. 'Beamten' [bə 'ʔampm] (*civil servants*). Fricatives have more distinctive acoustic structures separating the different places of articulation than nasals and stops, and among the latter, unreleased ones are

still less salient than released or even aspirated ones. Furthermore, the syllable or word initial position has a higher signalling value for a listener and must therefore be given a more precise articulation by a speaker. Thus the final position has a higher reduction coefficient than the initial one, allowing for instance final [bm] in 'geben' (*to give*), but preventing the sequence [pm] across the word boundaries in 'gib nicht' (*don't give*). What is not very distinctive for a listener anyway may be reduced by a speaker more easily to yield to the principle of economy of effort.

(b) Different communicative situations put different demands on the perceiver of speech, and speakers have to tune their performance to these conditions to guarantee a successful language interaction ([6]). This means that speaking styles in keeping with different speech environments exhibit varying degrees of reduction oriented towards the listener's needs

## 3. CONCLUSIONS

Instead of the dichotomy of phonetics and phonology or of multiple levels of phonetic representation I propose a unitary explanatory level, which I call "Phonetics in Language and Speech Communication". Besides this domain of an ontological status there may be as many heuristic ones as one deems necessary for preliminary descriptive orientation. In the pursuit of this communicative goal the study of word contrasts and lab speech has to be supplemented by the analysis of connected, continuous and spontaneous speech, which will make the investigation of articulatory reduction processes possible. They will in turn throw light on the organization of speech production and allow us to give a more precise account of the concept of articulatory effort in a language-independent perspective. They will also provide a direct link with historical sound changes. One aim must be the quantification of reduction coefficients for different speaking styles, replacing the atomisation of a multitude of articulatory parameters by global functional coordinative structures. They in turn need an auditory control unit that regulates the ease of articulation in relation to the perceptual demands of the communicative situation and incorporates language-independent perceptual distance measures.

## 4. REFERENCES

[1] "*Journal of Phonetics*", *18(3)* (1990).
[2] KOHLER, K. J. (1984), "Phonetic explanation in phonology: the feature fortis/lenis", *Phonetica*, *41*, 150-174.
[3] KOHLER, K. J. (1990), "Segmental reduction in connected speech in German: phonological facts and phonetic explanations"; in HARDCASTLE and MARCHAL (eds.) "*Speech production and speech modelling*", pp. 69-92, Dordrecht/Boston/London: Kluwer Academic Publishers.
[4] KOHLER, K. J. (1991), "The phonetics/phonology issue in the study of articulatory reduction", *Phonetica*, *48*.
[5] LADEFOGED, P. (1964), "*A Phonetic study of West African languages*", Cambridge: Cambridge University Press.
[6] LINDBLOM, B. (1990), "Explaining phonetic variation: a sketch of the H & H theory", in HARDCASTLE and MARCHAL (eds.) "*Speech production and speech modelling*", pp. 403-439, Dordrecht/Boston/London: Kluwer Academic Publishers.
[7] MILLER, G. A. and NICELY, F. E. (1955), "An analysis of perceptual confusions among some English consonants", *JASA*, *27*, 338-352.
[8] WESTERMANN, D. and WARD, I. C. (1933), "*Practical phonetics for students of African languages*", London: Oxford University Press.