Journal of Phonetics 39 (2011) 319–329

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Perception of phonetic detail in the identification of highly reduced words

Oliver Niebuhr [a,*], Klaus J. Kohler [b]

[a] Department of General and Comparative Linguistics, Christian-Albrechts-University, Kiel, Germany
[b] Institute of Phonetics and Digital Speech Processing, Christian-Albrechts-University, Kiel, Germany

ARTICLE INFO

Article history:
Received 16 January 2009
Received in revised form
1 December 2010
Accepted 10 December 2010
Available online 19 January 2011

ABSTRACT

There is great phonetic variation of words in context, conditioned by phonetic environment, word type, and speaking style in different communicative situations. Function words and modal particles are particularly susceptible to having their phonetic weight and complexity reduced, especially in casual spontaneous speech. But even if whole strings of segments are no longer delimitable in reduced forms compared with fuller pronunciations of the same lexical items, there will still be *articulatory prosodies*, superimposed upon the remaining sound material, which retain essential components of the fuller forms, the *phonetic essence* that characterizes the whole form class of a word. The extreme reduction [aĩĩ] of the German modal particle *eigentlich* 'actually' [aɪ(ɡ)ŋ.(t)(l)i(ç)] is a case in point. The length, palatality and nasality of its gliding movement reflect the polysyllabicity, the central nasal consonant and the final palatal syllable of the fuller forms. It is assumed that this phonetic essence triggers lexical identification in the listener. Therefore two perceptual identification experiments were carried out. They showed the crucial role of the duration of a palatal gliding section in the diphthong [aĩĩ] to distinguish between *eine*__ 'one__' and *eigentlich'ne*__ 'actually a__'. A third test showed further that listeners reacted differently to the palatal glide duration in different reduction environments, which may be related to different functional assessment of reduced forms in situational contexts.

1. Introduction

1.1. The analysis of reduced speech

Words take on various phonetic manifestations in connected speech, depending on word type, phonetic context, and speaking style, especially function words and modal particles. This is more extreme in spontaneous dialog. The traditional way of dealing with this phonetic variation is to group the variants around canonical forms, which represent the most elaborate citation form pronunciations, often guided by phonological transformation from orthographic form. This canonical representation is generally segmental phonemic, and the variants are derived from it by deletion, addition or modification of phonemic segments. If the changes, especially the modifications, cannot be mapped one-to-one onto phonemic representations, allophonic segmental statements are made at the phonetic level, e.g. when in *handbag* the realization of /ndb/ is neither [ndb] nor [mb] but [n͡mb] with coronal–labial double articulation, or when the assimilated fricative in *this* [ʃʃʲ] *shop* is different from the geminate fricative in *fish* [ʃʃˣ] *shop*. This segmental representation of words in relation to phonemic canonical forms is a useful sorting principle for pronunciation dictionaries, such as the reference works by Jones/Roach (1997) and Wells (1990) for English, and WDA—Wörterbuch der deutschen Aussprache (1969) for German.

The canonical-phonemic reference approach to phonetic variation of lexical entries is also a useful heuristic device for systematic descriptions of the pronunciation of a language, such as Gimson/Cruttenden (1962, 2008) for English, or Kohler (1995) for German, as well as for descriptive models of speech reduction, e.g. Kohler (1990, 1998, 2001) for German. Finally, the segmentation and labeling of acoustic speech corpora has greatly profited from this framework. However, it runs into conceptual problems when the distinctive features of vowels and consonants as well as their assimilation or elision are no longer linearly segmentable (cf. Gow, 2002; Heid & Hawkins, 2000; Local, 2003; Nolan, 1992), and when phoneme strings, which may extend beyond syllables to whole words, need to be marked as deleted qua segmental units although the signal portion is still recognized as containing the full lexical information in the utterance context.

As regards the latter case, the following example of highly reduced speech was found in the Kiel Corpus of Spontaneous Speech (IPDS, 1995, 1996): *ich kann Ihnen das ja mal sagen* 'I can

* Correspondence to: Freyastrasse 10, 24939 Flensburg, Kiel, Germany. Tel.: +49 461 940 3993.
E-mail addresses: niebuhr@linguistik.uni-kiel.de, oliver.niebuhr@t-online.de (O. Niebuhr).

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.12.003

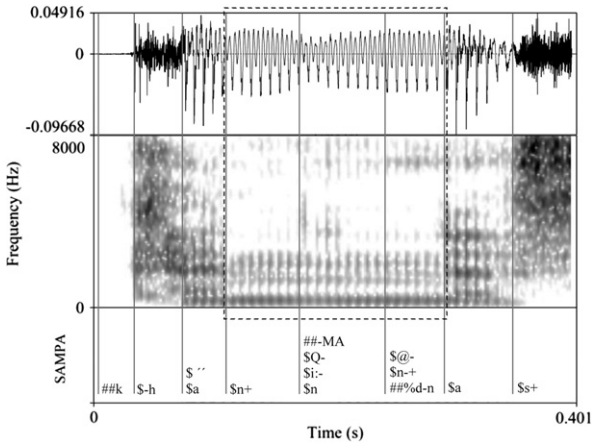

Fig. 1. Speech wave, spectrogram, and aligned SAMPA labels in the *Kiel Corpus* annotation system, for the spontaneous German utterance *ich kann Ihnen das ja mal sagen* 'I can mention this to you'. The dotted box rims the long alveolar nasal (ca. 180 ms) in between [ka___as].

mention this to you' (g072a015). Fig. 1 provides the speech wave, the spectrogram, as well as a linear segmentation and SAMPA labeling. There is a long stretch of an alveolar nasal (180 ms) in [ka___as] (cf. dotted box in Fig. 1), i.e. from the end of the word *kann* into the word *das*, in which there is initial nasalization instead of a lenis plosive. It is not possible to delimit the word *Ihnen*, whose citation form phonemic transcription is /iːnən/. Its usual pronunciation is without [ə] in the second syllable, but as is obvious from Fig. 1, the utterance shows no delimitable signal portion for a vowel [iː] either. However, the person who did the orthographic transliteration, and subsequently the labeller as well as the phoneticians who processed the data, had absolutely no doubt that the utterance contained *Ihnen*, and was not *ich kann das ja mal sagen*, without *Ihnen* but also with a long [n] in [ka___as], an equally possible utterance in the same semantic context of an appointment-making scenario. Looking more closely at the phonetic manifestation of the nasal stretch reveals that it has palatalization throughout, increasing towards the center. Kohler (1999) referred to such articulatory residues in the reduction of function words as *articulatory prosodies*, "which persist as non-linear, suprasegmental features of syllables, reflecting, e.g., nasality or labiality that is no longer tied to specific segmental units" (p. 89). Thus, articulatory prosodies are distinctive suprasegmental vocal-tract and phonation features that identify words in context in spite of segmental reduction. In the above example, the articulatory prosody of palatalitzation, in addition to the long duration of the signal portion of the alveolar nasal, which it overlays, references *Ihnen*.

Of course, what happens here from an articulatory point of view is that the tongue body articulation for the high front vowel [iː] in between the open central vowels of [ka___as] is not elided but is carried out while the tongue tip/blade forms contact with the alveolar ridge, resulting in palatalization rather than in an acoustic [iː] segment as defined by phoneticians. From this perspective, the weakly reduced form [iːn̪n̪] and the more strongly reduced form [nʲn̪ʲ] can be related to the same class (i.e. *Ihnen*) without an elaborate derivation from one canonical representation, because they both contain palatality and long alveolar nasality, as do other intermediate degrees of reduction. This means that all phonetic forms of this word must contain these features; they constitute the *phonetic essence* of *Ihnen*. This concept of phonetic essence may be assumed to apply to function words generally and possibly even to all lexical items. The phonetic essence of a lexical item manifests itself either in segmental units in the less reduced forms or as articulatory prosodies in more extreme reduction, where it appears to be sufficient for the listener to identify the word. Thus, [kanʲn̪ʲas], as against [kannas] *kann das*, can be decoded as containing *Ihnen* although there is no syntactic/semantic bias in the same linguistic and situational context.

Contrary to the conventions guiding vowel and consonant segmentation, such articulatory prosodies are not temporally delimited; all that matters is that they manifest themselves within a certain environment, where exactly can vary greatly. The extension may actually be quite large, as in the present case where palatality reaches beyond the nasal consonant in both directions: the aspirated plosive [k$^{h}_{+}$] of the preceding word *kann* is fronted, and its vowel as well as the one of the adjacent word *das* are raised and centralized. In *kann leider* of another utterance by the same speaker, where *kann* is also followed by an alveolar consonant and in turn by an open vowel element, this is not the case.

Glottalization is another articulatory prosody in German, used to signal interruption of modal voice in nasal context in place of a stop consonant. Thus, *könnten* 'could' may be differentiated from *können* 'can' [kœnn], either as [kœntn] or [kœn̰n], both creating the same kind of "break" in the acoustic signal for the listener. In the case of [t], the break is referable to a segmentally delimited stop occlusion inside a stretch of alveolar nasality; in the case of [n̰], it is superimposed on the nasal stretch and variable as to its extension and position. As long as a few irregular glottal pulses occur somewhere in the signal portion of [œnn] the listener decodes it as the lexical unit *könnten*, cf. Kohler (1999). The same articulatory prosody of glottalization can differentiate German *campen* 'to camp' [kʰɛm̰m] from *kämmen* 'to comb' [kʰɛmm] and '*sollten* 'should' [zɔl̰n̰] from *sollen* 'are to' [zɔln], for example, in the contexts *Zum campen/kämmen ist es noch zu früh* 'It is still too early to camp/comb' and *Sie sollten/sollen das machen* 'They should/are to do this'. Again, glottalization produces a signal break, as does the stop articulation of [t] and [p]. This break is part of the phonetic essence of such words as *könnten, campen* and *sollten*, in distinctive contrast to *können, kämmen* and *sollen*.

A third type of articulatory prosody in German is nasalitzation. Strongly reduced forms of *soll er* and *sollen wir* in the frame ___*das machen?* 'is he / are we to do this?' may be contrasted as [zɔɐ] versus [zɔmɐ] In *sollen wir*, the bilabial occlusion may not be reached but lowering of the velum still occurs, thus nasalization is superimposed on the vocalic resonances—[zɔ̃ɐ̃], replacing a segmentally delimitable nasal consonant by an articulatory prosody with the same function of signaling the phonetic essence of nasality to a listener, which is absent from *soll er*, cf. Kohler (1999). In other languages like English and French, observations which are interpretable in terms of articulatory prosodies concern, for example, long-term resonances of light and dark liquids (cf. Heid & Hawkins, 2000) or quality, f0 and duration changes of consonants in the context of schwa elision (cf. Gadet, 1992).

The segmentation and labeling conventions that were developed for the phonetic annotation of the Kiel Corpus took articulatory prosodies for lexical identification into account. If phonemic segments of a canonical base form of a lexical item cannot be mapped onto a signal portion corresponding to it they have to be marked as deleted, but if at the same time there is still an articulatory prosody that carries lexical identification it needs to be marked as a non-segmental insertion < -MA > (=prosodic marker) put before the segment string marked as deleted. So, *Ihnen das* in Fig. 1 is transcribed as

< ##-**MA** **SQ**- **Si**:- **Sn** **S@**- **n**-+ ##**%d-n** **Sa** **Ss**+ >

(for further details of the transcription system and conventions, cf. Kohler, Pätzold, & Simpson, 1995). Search operations on corpora annotated in this way can access the incidence of