Original Paper

Phonetica 2011;68:57–87
DOI: 10.1159/000328656

Received: March 16, 2011
Accepted: April 3, 2011

On the Role of Articulatory Prosodies in German Message Decoding

Klaus J. Kohler Oliver Niebuhr

University of Kiel, Kiel, Germany

Abstract

A theoretical framework for speech reduction is outlined in which 'coarticulation' and 'articulatory control' operate on sequences of 'opening-closing gestures' in linguistic and communicative settings, leading to suprasegmental properties – 'articulatory prosodies' – in the acoustic output. In linking this gestalt perspective in speech production to the role of phonetic detail in speech understanding, this paper reports on perception experiments that test listeners' reactions to varying extension of an 'articulatory prosody of palatality' in message identification. The point of departure for the experimental design was the German utterance *ich kann Ihnen das ja mal sagen* 'I can mention this to you' from the Kiel Corpus of Spontaneous Speech, which contains the palatalized stretch [k̟ʰɛ̈̆nʲnʲəs] for the sequence of function words /kan iːn(ə)n das/ *kann Ihnen das*. The utterance also makes sense without the personal pronoun *Ihnen*. Systematic experimental variation has shown that the extent of palatality has a highly significant influence on the decoding of *Ihnen* and that the effect of nasal consonant duration depends on the extension of palatality. These results are discussed in a plea to base future speech perception research on a paradigm that makes the traditional segment–prosody divide more permeable, and moves away from the generally practised phoneme orientation.

1. Modelling Speech Reduction

1.1. Coarticulation and Articulatory Control in Speech Production

In their seminal experimental investigation into speech articulation, Menzerath and de Lacerda [1933] showed that there is continuous movement of the articulators, which cannot be captured adequately by chains of segmental building blocks consisting of on-glide, hold and off-glide, as traditionally maintained by descriptive phonetics. The dynamic patterns are sequences of opening-closing movements, opening into a vocoid and closing into a contoid [in Pike's, 1943, terminology]. They are interwoven on the basis of two principles: 'synkinesis' – coarticulation in the proper sense of the

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/11/0682–0057$38.00/0
Accessible online at:
www.karger.com/pho

Prof. Dr. Klaus J. Kohler
Holm 4
DE–24113 Molfsee (Germany)
Tel. +49 431 651453, E-Mail kjk@ipds.uni-kiel.de

word, i.e. simultaneous movements of different articulators, and ‘articulatory control’ – control of the opening-closing movement of an articulator in relation to a focal point. The first type is illustrated by the closing movement [ut], where the tongue tip rises to alveolar closure during the lip and tongue dorsum articulations, the second type by the closing movement [ap], which is controlled by lip occlusion and ends in labial closure.

On the basis of these concepts of coarticulation and articulatory control, Öhman [1966] provided a more extensive and more systematic experimental data analysis of VCV syllables (C = voiced plosives) in Swedish and American English, this time in the acoustic domain of spectrographic measurement with much superior instrumental techniques, although the speech material was still highly stylized. He came to the following conclusions [Öhman, 1966, p. 165]:

> The data . . . suggest that in Swedish and English the variability of the formant transitions in VC sequences is controlled by the postconsonantal vowel . . . Since traces of the final vowel are observable already in the transition from the initial vowel to the consonant, it must be concluded that a motion toward the final vowel starts not much later than, or perhaps even simultaneously with, the onset of the stop-consonant gesture. A VCV utterance of the kind studied here can, accordingly, *not* be regarded as a linear sequence of three successive gestures. We have clear evidence that the stop consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture.

Thus contrary to the segment- and phoneme-based locus theory developed at Haskins, ‘the perception of the intervocalic stop must be based on an auditory analysis of the entire VCV pattern rather than on any constant formant frequency cue’ [Öhman, 1966, p. 165].

1.2. Phrase-Level Phonetics

1.2.1. The Control of Opening-Closing Gestures in Connected Speech

The concepts of coarticulation and articulatory control were applied to the investigation of connected speech processes in German read and spontaneous corpora, and further developed in a theory of phrase-level phonetics, modelling speech reduction, at the Institut für Phonetik und digitale Sprachverarbeitung (IPDS) Kiel over several decades [IPDS, 1994, 1995, 1996, 1997; Kohler, 1974, 1979, 1990, 2001a; Kohler et al., 1995; Wesener, 1999, 2001]. The theoretical stance can be summarized as follows:

Phrase-level speech phenomena are controlled by the principle of goal-oriented motor economy in the speaker, and checked by the need to maintain sufficient linguistic distinctivity for the listener depending on speaking style and communicative situation [Kohler, 1979; Lindblom, 1983, 1990]. Over and above paradigmatic differentiation of linguistic units, there is the need for syntagmatic structure in linguistic messages at a hierarchy of levels from syllables to words to morphological and syntactic constructions to semantic organization, and to prosodic grouping by accent, intonation and phrase boundaries. The prosodic features may support, or cut across, any of the former syntagmatic elements. These groupings are characterized by internal cohesion and junctural separation at the boundaries, signalled by segmental and prosodic indices [Kohler, 1983]. Internal cohesion raises the probability of phonetic fusion inside the various syntagmas, whereas their boundaries have a high probability of being marked by phonetic separators [Kohler, 1991]. The effect of syntagmatic cohesion on the degree of articulatory reduction may be demonstrated with an example from English. In *The patient's illness is the doctor's* bread and butter, the high internal cohesion of the idiomatic phrase allows reduction to [ˈbrɛb m ˈbʌtə], which would be far less likely in *We have to do some shopping, we need* bread and butter *and quite a few other things,* where the reduction stops at [ˈbrɛd n ˈbʌtə].

The constituents of phrase-level articulation are sequences of opening-closing movements of the vocal tract, which are basic speech gestures, ontogenetically and phylogenetically [Macneilage, 2008]. They are defined in their spatial and temporal dimensions with regard to the component articulators and cavities (oral, nasal, laryngeal). They are syllable-sized, i.e. 'larger than the sound segment', without being congruent with syllables. The term 'syllable' refers to the phonological category, the term 'gesture' to the unit of opening-closing movement in speech production. For example, English or German *text* is one syllable but two opening-closing gestures. Speech analysis on this theoretical basis transcends the atomic segmental switches of alphabetic transcriptions and looks at the production, modification and reorganization of whole syllable-sized units. It is the whole opening-closing gesture that is affected at the phrase level, not the isolated segments that descriptive phonetics may excerpt from this more global unit.

These principles of phrase-level phonetics will be discussed in some detail with reference to connected speech phenomena that have been found in the German read and spontaneous corpora. The discussion starts with a selection of simple opening-closing gestures and progresses through their articulatory reorganization under increasingly complex contextual conditions to the formation of long components of glottalization, nasality, and palatality. It finally leads to the analysis of a long component of palatality in a spontaneous speech example that forms the basis for an exemplary testing of the relevance of these long components for message decoding.

Let us first examine the manifestations of post-stress /Cən/ syllables. In South German dialectal varieties, the tongue-tip closing movement is generally not carried out if there is not a subsequent unstressed vocoid opening in the same lexical unit, exemplified by *ebe(n)* versus *ebene* 'even (adj.)'. In North German standard varieties, on the other hand, an oral closing movement remains, and the opening movement may be curtailed instead, in what way, depends on the gesture onset. If C = [l], the tongue-tip closing movement may directly carry on to complete oral closure from the initial lateral configuration, and the velum is lowered somewhere in the course of the complete gestural unit of the tongue tip, e.g. *stellen* [ˈʃtɛln] 'put', just like *Köln.* If C = [ʁ], the complete opening-closing gesture may be integrated into the preceding gesture, reorganizing its vocoid opening, e.g. *fahren* [ˈfaːʁən] > [ˈfɑːn] 'drive', *studieren* [ʃtʊˈdiːʁən] > [ʃtʊˈdiːɐn]. If C = [n], the elimination of an opening movement results in a long tongue-tip closure, as in *kennen* [ˈkɛnː] 'know'. If C = [m, ŋ], the labial or dorsal onset may control the entire gesture and lead to an elimination of a tongue-tip movement, as in *kommen* [ˈkɔmː] 'come', *fangen* [ˈfaŋː] 'catch'.

If C = [t, d], the opening gesture may be effected by nasal plosion instead of an oral release, as in *reden* [ˈreːd^{n}n] 'talk', *raten* [ˈraːt^{n}n] 'guess'. If C = [b, p, g, k], the gesture may be reorganized in such a way that the tongue-tip movement is eliminated altogether and the labial or velar closure onset controls the place target of the offset, the opening gesture again being effected by nasal plosion, as in *leben* [ˈleːb^{m}m] 'live', *Lappen* [ˈlap^{m}m] 'cloth', *legen* [ˈleːg^{ŋ}ŋ] 'lay', *lecken* [ˈlɛk^{ŋ}ŋ] 'lick'. In the case of the much shorter occlusion of the lenis plosives, the timing of velic lowering may lead to the elimination of an oral stop phase.

If C = [s, ʃ], a curtailing of the vocoid opening does not simply entail a continuation of the tongue-tip movement to complete closure, as for *stellen*, this time from a fricative stricture, but requires the much more complex, quite precise coordination of airflow, glottal adduction and velic lowering, in e.g. *lassen* [ˈlasn] 'let', *waschen* [ˈvaʃn] 'wash'. This greater complexity remains if, for C = [f, ç, x], the labial or dorsal onset controls the closing gesture and the tongue-tip movement is eliminated. Contrary to what happens in the opening-closing gestures with plosive onsets, place control and elimination of the tongue-tip movement do not simplify the gestures in the case of fricative onsets. Consequently, coarticulation of the tongue-tip movement with the labial or dorsal stricture is the most frequent manifestation of this type of gesture in, e.g., *rufen* [ˈʁuːfn] 'call', *streichen* [ˈʃtʁ̥aɪçn] 'delete', *machen* [ˈmaxn] 'make'. For [ɱ, ɲ, ŋ] to occur, more contextual conditioning is required in the subsequent opening-closing gesture of the next word, for instance in *Rufen wir ihn doch an* [fɱv]. 'Let's call him.', *Die streichen ja alle von der Liste.* [çɲj] 'They take everybody off the list.', *Die machen gar nichts*. [xŋg] 'They do not do anything at all.'

The lack of a vocoid opening in the gestural types discussed so far is standard in modern North German, its presence is reinforcement, rather than the former being a reduction of the latter. The cases of articulatory control of place and nasality, however, presuppose a less formal speech style and communicative situation. When the opening-closing gesture preceding [{ptk}ən] ends in a nasal stop [m, n, ŋ], the reorganization can go still further, especially in more relaxed spontaneous speech: the velum stays in a lowered position from this focal nasal point right to the end of the next gestural unit, and the occurrence of an articulatory stoppage at the juncture between the two gestures is signalled by a glottal break in modal nasal voice, either a glottal stop or glottalization. The simplest case of a long tongue-tip occlusion is illustrated by two realizations of *könnten* [ˈkœnn̰n] 'could' in figure 1.

In figure 1a, glottalization is superimposed on the nasal consonant at its centre; in figure 1b, there is modal-voice nasalization in the vowel, which is immediately followed by a glottalized nasal contoid that finally turns into modal voice again. Examples of *könnten* in the Kiel Corpus show great fluctuation in the positioning of the glottalized period across the stretch of utterance corresponding to /œntn/ , cf. Kohler [2001b] and audio examples at the URL http://www.ipds.uni-kiel.de/kjk/pub_exx/kongrbtr/plosglot.html. This temporal flexibility of glottalization in its synchronization with vocal-tract dynamics shows that the precise point of occurrence and temporal extension in the nasal stretch are unimportant as long as a perceptible break in modal voice is created for the listener to differentiate the utterance from *können* [ˈkœnn] 'can' [Kohler, 1999, 2001b].

In the example *zwischen Montag dem siebenten und Freitag dem elften Februar wäre mir recht* 'between Monday the seventh and Friday the eleventh of February would be fine for me' (HAH g076a010) from the Kiel Corpus of Spontaneous Speech, the more complex gestural sequence [mtn] of *siebenten* is also realized as a glottal break in continuous nasality [ˈziːm͡n̰n]. But here a labial closure is coarticulated with a tongue-tip movement, which opens into the vocoid of the next gesture, the conjunction *und* [ʊn] concatenating two appointment dates in a cohesive syntagmatic structure (fig. 2a). The opening phase of a gestural unit receives greater articulatory and perceptual weight than the closing phase. This can account for the absence of labial control across the entire gesture. If *siebenten* is followed by *März* or *Mai* in close syntagmatic cohesion the probability of the long labial occlusion [ˈziːmm̰m] is much higher, now triggered in look-ahead control [Kohler, 1976], as also happens in *elften Februar*

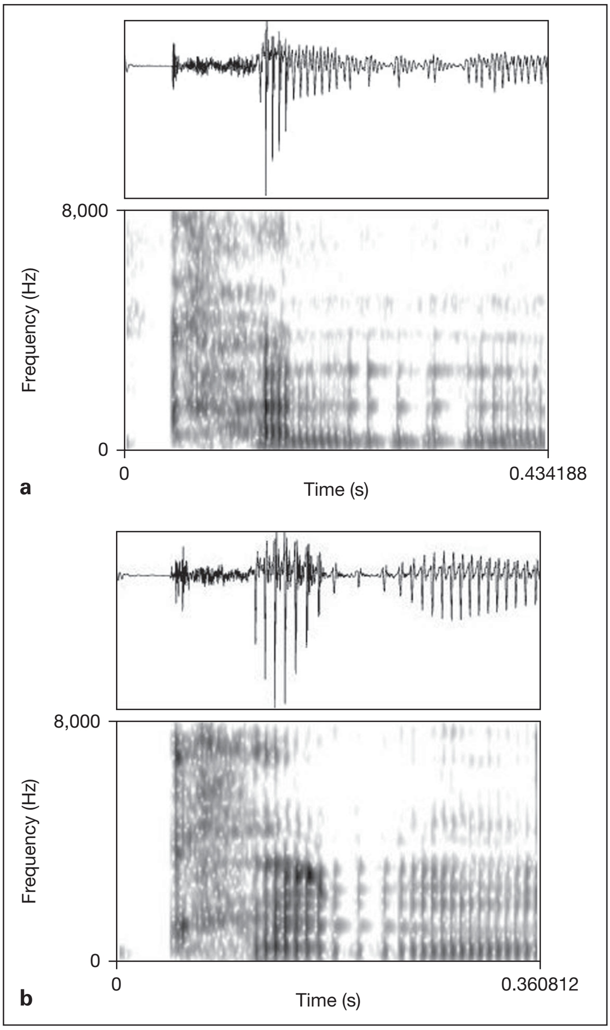

Fig. 1. Speech waves and spectrograms of *könnten* ‘could’ taken from the spontaneous German utterances KAE g086a004 (**a**) and HAH g074a000 (**b**).

[ˈʔɛ̰lfɱ̰ feːbʁuɑ] (fig. 2b), and as it can occur in *sie könnten mir/können mir vielleicht helfen* [ˈkœmɱ̰ mɪɐ]/[ˈkœm mɪɐ] ‘they could/can perhaps help me’. On the other hand, in another example from the Kiel Corpus of Spontaneous Speech, *wie wär’s mit dem vierten und siebenten?* ‘what about the fourth and eleventh’ (AME g312a003), the nasal gesture occurs utterance-final and does show labial control (fig. 3). The potential occurrence of articulatory control in the closing phase of a gesture and its obligatory absence in the opening phase is illustrated by *die Beamten* [bəˈa̰mm̰m] versus *Beamte* [bəˈa̰mtʰə] ‘civil servants’ [Kohler, 1992].

In the examples discussed above, we are dealing with suprasegmental nasality and glottalization, which are no longer tied to specific segments, but are superimposed on global gestural units: they are distinctive ‘articulatory prosodies’ [Kohler, 1999]. Further

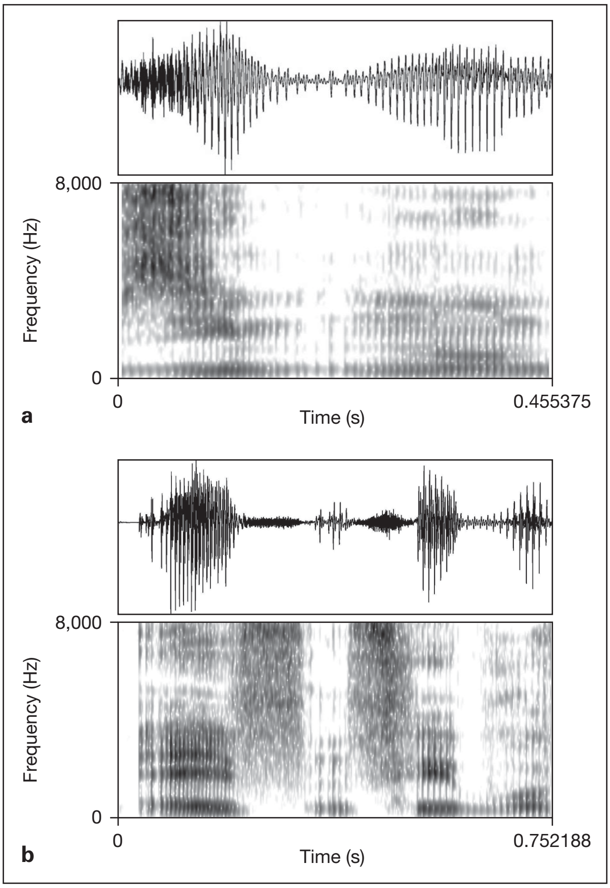

Fig. 2. Speech waves and spectrograms of the spontaneous German utterances *siebenten und* [ˈziːm͡n̰n ʊn] 'seventh and' (**a**) and *elften Februar* [ˈʔɛ̰lfm̰ feːbʁṵɑ] 'eleventh of February' (**b**); HAH g076a010.

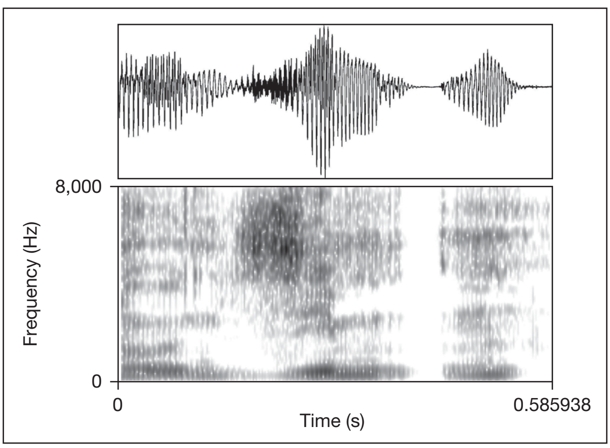

Fig. 3. Speech wave and spectrogram of the spontaneous German utterance *und siebenten* [ʊn ˈziːmm̰ʔm] 'and seventh'; AME g312a003.

examples of articulatory prosodies in German connected speech production, among others of glottalization, nasalization, velarization, labio(dent)alization, can be found in Kohler [1998] and Wesener [2001], with audio illustrations at the following URL:

http://www.ipds.uni-kiel.de/kjk/pub_exx/kk1998_1/kk_98a.html

http://www.ipds.uni-kiel.de/kjk/pub_exx/tw2001_1/hoerbsp-tw.html

1.2.2. Reduction of Function Words

Basically the same articulatory control processes, as described in the previous section, also occur in function words. But on account of their intrinsically low semantics, the articulation of function words is reorganized and reduced to a particularly high degree in connected speech if they are not highlighted by prosodic means for communicative functions, such as accentuation for emphasis of contrast. Function words are thus adapted more frequently and more extremely in fine gradation to the phonetic, linguistic and situational context. For example, the German sequence *mit dem*, preposition + inflected definite article 'with the', as in *mit dem Auto* 'by car' shows the following phonetic exponency: [mɪd̥em], [mɪd̥əm], [mɪpm], [mɪm̰m], [mɪm]. In *mit Demokraten* 'with democrats', reduction cannot go further than [mɪ d̥emoˈkʁ̥aːtn], and in *mitten* 'in the middle', it stops at [mɪtn]. When *guten* 'good (inflected)' has the full lexical semantics, as in *guten Appetit*, it is [ɡuːtn], but as part of a greeting, e.g. in *guten Abend* 'good evening', its semantic content may get weakened and reduction can go to [ɡʊdn], [ɡʊn], [n].

Research into speech reduction has been taking note of these special conditions controlling the exponency of function words in the phonetic descriptions of a variety of languages. Daniel Jones's [1956] *weak* and *strong forms* are an early account in the description of English, more particularly Received Pronunciation. Kohler [1979, 1990] gave a rule-based report on German, which was supplemented by an assessment of data from the Kiel Corpus of Read and Spontaneous Speech [IPDS, 1994, 1995, 1996, 1997] in Kohler [2001a]. In both the English and the German descriptions, the reduction of function words was linked to their default occurrence in deaccented sentence position. But the phenomenon was also recorded for French, a language without lexical stress and without the Germanic accent system, by Passy [1890, 1929; for further details cf. Kohler, 2002].

Extreme articulatory control of opening-closing gestures in function words may be illustrated by the phonetic exponency of two examples from the Kiel Corpus of Spontaneous Speech. First, the word *eigentlich*, which is either a content word meaning 'in reality', or a modal particle with the non-specific meaning of a filler 'really'. In its former use, an elaborate citation-form pronunciation can be [ˈaɪɡəntlɪç], but it was not found among the frequent occurrences of the word in the Kiel Corpus [Kohler, 2001a, 2008]. Figure 4 shows three stages of reduction in the two opening-closing gestures [ɡəntlɪç] of the modal particle. The first gesture can be levelled to [ɡŋ] and further to [ŋː], as described in 1.2.1. The second gesture, the unstressed suffix syllable [lɪç], is characterized by palatality throughout, i.e. by a high elevation of the tongue dorsum in [ɪç], as well as in the clear (palatalized) [lʲ]. In [ɪç] the vowel is, moreover, produced with a higher tongue position than before non-palatal consonants, e.g. in the suffix *'-nis'*. So the difference between vowel and fricative in this syllable is one between vibrating and open glottis with very similar tongue height. Lack of stress reduces the airflow and thus the generation of local friction. Furthermore, palatalized lateral and high front vowel are articulatory opposites in

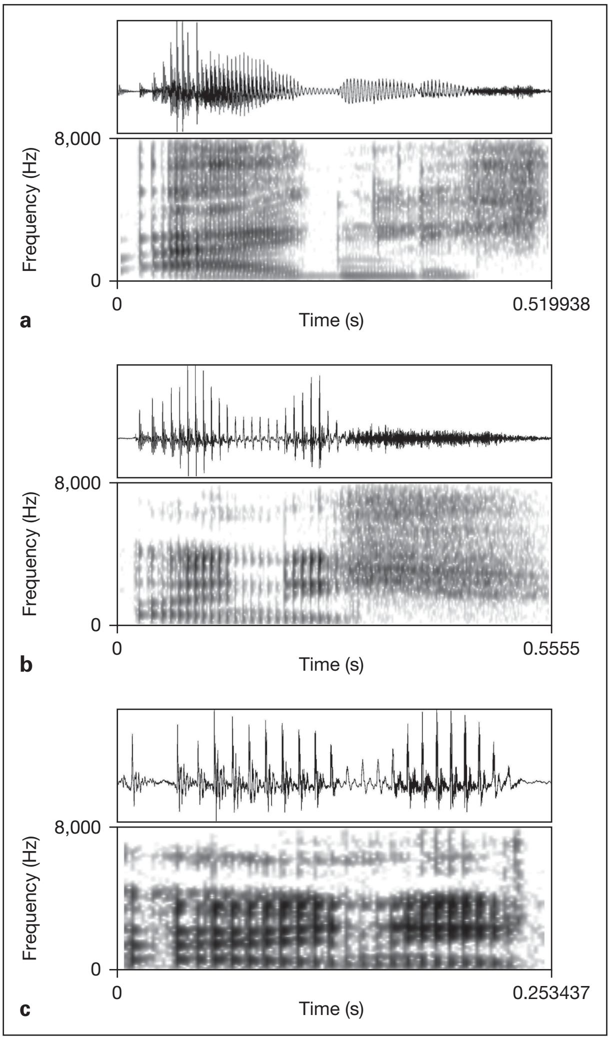

Fig. 4. Speech waves and spectrograms of three occurrences of *eigentlich* from the German Kiel Corpus of Spontaneous Speech: [ˈaɪgŋlɪç] FRS g092a021 (**a**), [ˈaɪŋ̟ːiç] MAW g427a005 (**b**), [ˈaɪŋ̟i] MAW g427a001 (**c**).

their central and side tongue-palate contacts, which puts high demands on the execution of the speech gesture chaining. In unstressed position, this can result in the elimination of the tongue-tip movement at the junction of the two gestures and in their integration into a continuous dorsal raising (cp. *fangen* [ˈfaŋː] in 1.2.1). The result of this articulatory control is [ˈaɪŋ̟i].

There is no record in the Kiel Corpus of Spontaneous Speech of gestural reorganization in *eigentlich* going beyond [ˈaɪŋ̟i], but it is possible resulting in one continuous dorsal closing gesture with superimposed velic opening [ãĩj̃], for instance in the phrase *das ist eigentlich ganz guter Wein* ‘this is really quite good wine’. It contains the essential components of the fuller forms, namely extended palatal-dorsal movement with interspersed nasality [Niebuhr and Kohler, in press]. This can lead to a potential

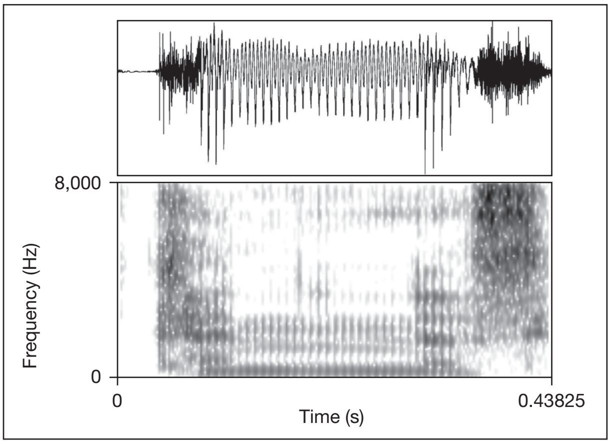

Fig. 5. Speech wave and spectrogram of *kann Ihnen das*, taken from the spontaneous German utterance *(ich) kann Ihnen das ja mal sagen)* ‘I can mention this to you’, TIS g072a015.

contrast between a long palatal and a velar articulatory prosody in *das ist* ***eigentlich*** *ganz guter Wein* [ãĩj̃(ɲ)] versus *das ist* ***ein*** *ganz guter Wein* [ãĩŋ] ‘this is (really) quite a good wine’. The elimination of a nasal stop occlusion in a sequence of function words and the gestural integration into continuous dorsal and labial movements with simultaneous velic lowering is documented in another example from the Kiel Corpus of Spontaneous Speech, ***nun wollen wir mal*** *kucken* [nʊ̞̃ ɔ̞̃ːnᵛ ɐ̟̃ ɱa] ‘now let’s see’ (OLV g122a009) [Kohler, 2000].

The second illustration of extreme articulatory control in function words concerns the sequence of function words *kann Ihnen das* in the sentence frame *ich__ja mal sagen* ‘I can mention this to you’ from the Kiel Corpus of Spontaneous Speech turn g072a015 (fig. 5). Its elaborate citation form pronunciation is [kan iːnən das]. The third opening-closing can again be levelled as in *kennen* [ˈkɛnː] of 1.2.1. But here the articulatory control of the opening-closing gestures can also go further in two respects. Velic raising for the tongue-tip occlusion at the beginning of the last function word may be delayed until the vocoid, which is in turn reduced to [ə], and the opening into the vocoid of the second gesture may not take place at all, thus creating a long nasal tongue-tip closure. Of course, the tongue body articulation for the high front vowel [iː] in between the vocoids of [ka__əs] is carried out while the tongue tip forms contact with the alveolar ridge for the long nasal, thus resulting in high front vowel resonance during its articulation, i.e. the secondary articulation of palatalization, rather than an acoustic [iː] segment. The spectrogram also shows that the degree of palatalization through tongue elevation increases in the centre of the nasal stretch. Again an adequate representation of this utterance needs to take this long component of palatalization into account as an essential ingredient in its production, an ‘articulatory prosody of palatality’[1] over and above segmental units.

Contrary to the conventions guiding vowel and consonant segmentation, such articulatory prosodies are not temporally delimited; they manifest themselves within a

[1] In this paper, the terms palatalization and palatalized are used with reference to the speech production category of secondary articulation in consonants [IPA, 1999, p. 17] and, by extension, to anticipatory and carry-over effects on tongue position in vowels and consonants surrounding such secondary articulation; the term palatality, for example in the collocation with articulatory prosody, refers to the total of speech signal properties that contribute to a perceptual feature acute, as against grave, in the differentiation of long stretches of utterance.

certain environment, where exactly can vary greatly. In the above example, the vowel of *kann* is raised and centralized, when compared with other occurrences of the word from the same speaker in the Kiel Corpus. This articulatory prosody of palatality differentiates the utterance from the equally possible one without it, *ich **kann das** ja mal sagen* [k^{h}annas], which does not contain *Ihnen.*

1.2.3. Phrase-Level Phonetics and Articulatory Phonology

The concept of speech gestures outlined in the preceding sections as part of a theory of phrase-level phonetics in speech communication differs in important points from that of Articulatory Phonology [Browman and Goldstein, 1992; Kohler, 1992, 2001c].

The theory of phrase-level phonetics moves a good deal further away from a linear segmental phonemic framework because the primary gestures are considered to be syllable-size units and articulatory variability is construed as a reorganization of these global dynamic structures according to internal and external conditions. Segment-size units are secondary and result from segmentation of the global gestures. This contrasts with the gestural score in Articulatory Phonology, which is based on successive phonological segments as the primary elements whose coordinated gestural parameter specifications are temporally and dynamically variable in concatenation.

Syllable-size gestures incorporate segments and long componential features, as units in their own right. Articulatory Phonology only recognizes segments, feature spreading is a consequence of segmental gestural sequencing. In the theory of phrase-level phonetics, gestural interaction applies to global structures of flexible extension and with a high degree of internal cohesion, in Articulatory Phonology the interaction is local between juxtaposed segments. This basic approach has not been changed by the introduction of the concepts of a phase window and of prosodic boundary or π-gestures in the development of Articulatory Phonology by Byrd [1996] and Byrd et al. [2000].

In phrase-level phonetics, components of opening-closing gestures may disappear or be changed to others; in Articulatory Phonology gestures that are incorporated in the gestural score cannot disappear nor be changed to other gestures. Syllable-size gestures are not articulations that are deaf to the auditory consequences, as is the case for segment-size gestures in Articulatory Phonology. Syllable-size gestures are embedded in communicative functions which determine their realization, whereas Articulatory Phonology does not incorporate the functional aspect and treats gestures as mechanical processes without cognitive links.

As Articulatory Phonology only recognizes overlap and magnitude of juxtapositional segment-oriented gestures as the sources of phonetic variability in the execution of a constant phonological gestural score there is no room for the distinction between coarticulation and articulatory control. Even if there are no observable traces of velic raising-lowering, of glottal abduction-adduction and of tongue-tip movement in /ntnm/ of *könnten mir* (see 1.2.1), the gestures specified by these tract variables are still supposed to be there because they belong to the phonological score, they are simply levelled by temporal sliding and reduction to zero magnitude in contiguous gesture concatenation. But this stance gets into unsolvable difficulties in view of such phonetic forms as [ˈkœmm̰ mɪɐ], where labiality is initiated by a labial gesture several removes from the one currently being executed, and where the glottal gesture is not abduction as in the elaborate realization [ˈkœnt$^{\text{h}}$ən mɪɐ] but the exact opposite, i.e. compression resulting in glottalization or glottal closure, and the absence of velic raising-lowering is more economically and more convincingly modelled as the removal of a closing-opening gesture in the bilateral environment of long velic lowering. Such phenomena of speech reduction need to be captured by a concept of dynamic articulatory control by the side of local coarticulation. It does not make sense to refer [ˈkœmm̰ mɪɐ] to 'coarticulation' on the basis of gesturally defined overlap of phonological segments.

1.2.4. Articulatory Prosodies and Phonetic Essence

Relating the phonetic variability of lexical items to the dynamics of sequences of opening-closing gestures under the principles of coarticulation and articulatory control in meaningful linguistic and communicative frames leads, in direct derivation, to the recognition of persisting articulatory components, termed 'articulatory prosodies' in 1.2.1, following Kohler [1994, 1998, 1999]. They constitute the 'glue' that gives internal coherence to each lexical exponent and provides the essential articulatory characteristic common to all the phonetic exponents of a lexical item, termed their 'phonetic essence' in Niebuhr and Kohler [in press]. Articulatory prosodies contribute to the identity of a lexical item across its varying manifestations, whereas post hoc linear phonemic segmentation stresses divergences from canonical forms via deletion, assimilation and insertion.

Since Firth's [1948] paper, the phonological relevance of fine phonetic detail beyond segmental phonemic representation has been studied in a fair number of investigations [cf. among others Hawkins and Nguyen, 2004; Kelly and Local, 1989; Local, 2003; Simpson, 1992; West, 2000]. A good proportion are descriptive rather than experimental [Kelly and Local, 1989, refer to long-domain resonance patterns on an auditory descriptive basis], and perceptual experimental analyses are in the minority and have so far not been based on stimulus generation from spontaneously produced utterances, nor have they tested the relevance of long articulatory components for the perception of reduced speech. These new aspects are at the centre of the experiments reported in this paper.

Hawkins and Smith [2001] and Hawkins [2003] have developed a theoretical framework for the role of fine phonetic detail in speech understanding. This paper follows the same reasoning that subtle aspects of vocal-tract dynamics in natural speech provide the output with acoustic coherence for auditory processing, and phonetic complexity beyond postulated segmental phonemic feature bundles is thus of fundamental importance for the listener. Under this perspective, articulatory prosodies superimposed on remaining sound material in speech reduction retain the phonetic essence of the whole class of phonetic manifestations of a lexical item and are thus directly relevant for message identification in speech perception. It is therefore not necessary for the listener to reconstruct canonical forms through phonemic restoration and top-down semantic interpretation, as implied in the experimental analyses by Ernestus et al. [2002] and Kemps et al. [2004; cf. the critique in Niebuhr and Kohler, in press]. This strips the phoneme concept of its widely held perceptual status, and relegates it to a useful heuristic device for alphabetic representation of speech in a variety of linguistic operations.

Conceptualizing the role of persisting phonetic detail across formal variability in speech understanding is intimately tied to modelling speech reduction as coarticulation and motor control in phrase-level phonetics. So, the discovery of articulatory prosodies in spontaneous speech data prompts the question as to how these suprasegmental phonetic properties are mapped onto perception and cognitive processing of utterances.

Kohler [1999] looked into the effects of nasalization and glottalization for the differential decoding of ***soll er/sollen wir*** *das machen?* [zɔ ɐ]/[zɔ̃ ɐ] 'is he/are we to do it?' and *die* ***können/könnten*** *uns abholen.* [kʰœnn]/[kʰœn̰n] 'they can/could collect us' in a series of perception experiments. (For audio illustrations see http://www.ipds.uni-kiel.de/kjk/pub_exx/kk1999_1/kk_99a.html). The results indicate that the presence or absence of nasality in the first pair of function words triggers the identification of one or the other meaning of the utterance, and that the presence of glottalization in a stretch of nasal resonance, irrespective of its extension and position distinguishes utterances containing *könnten* as against *können*. Similarly, Niebuhr and Kohler [in press] showed the importance of long palatality in *eigentlich 'ne rote* versus *eine rote* [aĩĩnʲə]/[aĩnə] 'actually a red one'/'a single red one' for semantic differentiation. The experiment reported in this paper continues this investigation into the role of the articulatory prosody of palatality in German message decoding.

2. The Decoding of Reduced Speech

2.1. The Research Questions

The example *ich kann Ihnen das ja mal sagen* 'I can mention this to you' from the Kiel Corpus of Spontaneous Speech, introduced in 1.2.2, is taken as point of departure for an investigation into the influence of varying degrees of an articulatory prosody of palatality on the perception and cognitive processing of utterances. The phrase was spoken as an aside in the following appointment-making context: *Wo ich im Juni Zeit hätte – ich kann Ihnen das ja mal sagen – wäre* . . . 'When I would be free in June – I can mention this to you – would be . . .' In this situational context, speech production is low-key in its laryngeal and vocal-tract parameters:

- The pitch level is lowered and smoothed, only the final verb receives a pitch accent in the form of a continuation rise leading back to the main utterance, none of the other words are accented.
- Vocal-tract movements are narrowed (a) by eliminating an opening-closing tongue-tip gesture for /niːn/ in *kann Ihnen* and executing the tongue-body raising during the long nasal consonant, (b) by lowering the velum in the initial vowel of the utterance section *ann Ihnen da* and not raising it again until its final vowel, (c) by shortening the /a/ vowels and curtailing their opening, thus fitting them into the overall reduced movement pattern, and (d) by adjusting the aspiration part of the /k/ release to the subsequent raised and centralized vowel.

This is the situational and phonetic environment that induces extreme reductions of function words, as described for this spontaneous speech example. We ask three interrelated questions regarding the identification of this utterance or of a corresponding one not containing *Ihnen*, in the same context (cf. 1.2.2):

(1) How is identification influenced (a) by the extent of palatality across the long nasal resonance, (b) by the extended palatality in the fronted and centralized vowel and in the adjusted release friction of the preceding [k̟ʰɛ̈], and (c) by the raising and centralization of the vowel in the following [əs]?

(2) How is it influenced by the duration of the nasal consonant in the different environments of (1) (a–c) considering that *Ihnen* contains a nasal in addition to the nasal of *kann*?

(3) How is it influenced by changes in the prominence of *kann*? The f0 pattern in this word affects its prominence, which may, in turn, have an effect on the strength of the nasal duration cue for *Ihnen*. Can we find an interplay between word and prominence perception, particularly when the palatality features (a) and (b) in (1) have been removed?

2.2. Hypotheses

The following hypotheses ensue from the research questions in 2.1.

Hypothesis 1 – Effect of Palatality. Decoding the utterance as containing *Ihnen* decreases with the successive removal of palatality from its extension across the section [k̟ʰɛ̈nʲnʲ] in the original utterance

(a) by cutting out the central part of increased palatalization in the nasal,

(b) by replacing [k̟ʰɛ̈nʲnʲ] altogether with non-palatalized [kʰann] from another *kann* in a non-palatal context by the same speaker;

(c) it decreases even more when the internal tie of the whole phrase *kann Ihnen das* is further loosened by additionally replacing the raised and centralized vowel of [əs] in *das* with the vowel of the non-palatalized [kʰann].

Hypothesis 2 – Effect of Nasal Duration. Kiel Corpus examples by the same speaker show that nasals are substantially longer when they contain [nʲnʲ] *Ihnen*, after a word ending in /n/ and before [əs] *das*, than when they do not. It is therefore to be expected that the perception of reduced *kann Ihnen das* versus *kann das* will be influenced by this duration factor as well. Decoding the utterance as containing *Ihnen* will decrease with the shortening of the nasal consonant duration within the different frames of palatality.

Hypothesis 3 – Effect of Prominence. Increased prominence of *kann* due to changed f0 pattern affects the link of nasal consonant duration to *Ihnen*. Consequently, in decoding the utterance as containing *Ihnen*, the cue value of the nasal duration variable is differentially influenced by the f0 pattern in *kann*.

3. Method

3.1. Properties of Selected Data from the Kiel Corpus

The original utterance *ich kann Ihnen das ja mal sagen* was further analysed in its acoustic parameters and compared with the analysis of other Kiel Corpus examples containing the words *kann*, *Ihnen, das* in order to create a basis for the generation of different extensions of palatality and nasal consonant durations in a systematic experimental design.

The excerpt *kann Ihnen das* [k̟ʰɛ̈nʲnʲəs] has strong, increasing – decreasing palatalization in a very long [n] of 180 ms duration from the offset of /ka/ to the onset of /as/, representing *Ihnen*. The dorsal plosive, its release burst and its immediately following local friction in *kann* are fronted, no doubt under the additional influence of the preceding palatal syllable *ich* [iç], and the aspiration noise immediately preceding the vowel matches the vowel spectrum. The spectrogram in figure 6a shows a concentration of energy in the aspiration noise in [k̟ʰɛ̈] around F2 and F3 of the vowel: the 2nd and 3rd spectral peaks, 12 ms before vowel onset, at 1,782 and 2,518 Hz, tie in with the F2 and F3 frequencies in mid-vowel. The vowels in *kann* [k̟ʰɛ̈nʲ] and *das* [nʲəs] are short as well as raised and centralized: [ɛ̈] – 33 ms, mid-vowel F1 = 575 Hz, F2 = 1,770 Hz, F3 = 2,617 Hz; [ə] – 44 ms, mid-vowel F1 = 496 Hz, F2 = 1,554 Hz, F3 = 2,429 Hz.

The same speaker produced a comparable case of vowel raising before *Ihnen* and reduction to a schwa vowel in following *das* in the utterance *wenn Ihnen das recht ist* ‘if that’s ok with you’ (g072a017) [ven iːnʲnʲəs]: [e] – 43 ms, mid-vowel F1 = 430 Hz, F2 = 2,032 Hz, F3 = 2,503 Hz; [ə] – 45 ms, mid-vowel F1 = 493 Hz, F2 = 1,536 Hz, F3 = 2,438 Hz; [nʲnʲ] – 113 ms. This can be compared with another utterance of his, *wenn das* [vɛn nəs] *klappen würde* ‘if that could be arranged’ (g073a002), without *Ihnen*, where the long nasal consonant is not palatalized, the vowel in *wenn* is more open and *das* again has a schwa vowel; the durations and formant frequencies of the vowels are [ɛ] – 48 ms, mid-vowel F1 = 542 Hz, F2 = 1,611 Hz, F3 = 2,365; [ə] – 36 ms, mid-vowel F1 = 587 Hz, F2 = 1501 Hz, F3 = 2,450 Hz. The fairly long duration of 75 ms for the nasal consonant in this unstressed intervocalic function word position indicates that the utterance is *wenn das*, not *wenn es* (i.e. ‘if that’, not ‘if it’), even with a reduced vowel [ə].

These examples show that the palatalization of the long nasal consonant, the main residue of *Ihnen*, also triggers an anticipatory extension into the preceding vowel, creating an articulatory prosody of palatality over a long stretch of utterance. In all the occurrences of *das*, irrespective of *Ihnen* preceding, the centralization of the vowel goes further. This is no doubt primarily the consequence of the function word being completely deaccented in all these cases. But in [k̟ʰɛ̈nʲnʲəs] the schwa vowel is also a better fit following on from an articulatory prosody of palatality than a low non-centralized vowel. So, there is justification in treating it as a carry-over extension of palatality, also contributing to the phonetic ‘glue’ of the whole phrase, into which *Ihnen* is integrated.

The same speaker also provides an instance of *kann* that is not followed by *Ihnen*, in the context *ich kann ˈleider also die ˈerste Zeit überˈhaupt ˌnicht* ‘well, unfortunately I can’t manage during the first period at all’ (g075a000; primary and secondary accents marked by ˈ and ˌ); *ich* has an extremely short vowel [i] and a very long fricative [ç], signalling a hesitation, but the rest of the utterance is fluent. Here *kann* is not a function word but a content word, meaning ‘can manage’, yet deaccented. The release burst and local friction of the dorsal plosive are again fronted under the influence of the preceding palatal fricative. However, the vowel is more open than in *kann Ihnen* of g072a015 and may be transcribed as [a] with mid-vowel F1 = 728 Hz, F2 = 1,584 Hz, F3 = 2,201 Hz. The aspiration noise immediately preceding it shows a concentration of energy in a lower part of the spectrum around the lower F2 and F3 frequencies of the vowel: the 2nd and 3rd spectral peaks, 12 ms before vowel onset, at 1,674 and 2,280 Hz, tie in with the F2 and F3 frequencies in mid-vowel (cf. also the spectrogram in fig. 6b). The vowel has a duration of 60 ms and is thus substantially longer than the vowel in *kann* of g072a015, where deaccentuation occurs in a function word and, at the same time, in an aside. The nasal consonant is not palatalized and has a duration of 110 ms, which is quite long and mirrors the status of a content rather than a function word. Another instance of *kann* in a non-palatal context occurs in the Kiel Corpus example *na ja, da kann ich überˈhaupt ˌnicht* ‘well now, I can’t manage then at all’ (g072a009) by the same speaker. It is again uttered in a low-pitched aside, but this time as a completely unaccented content word. Here the nasal consonant is short, only measuring 58 ms in intervocalic position.

3.2. Stimulus Generation

The generation of test stimuli builds on this data analysis. The utterance g072a015 *ich kann Ihnen das ja mal sagen* provides the base (token 1), which is subsequently manipulated to generate further test tokens. The excerpt *kann* from g075a000 (*ich kann ˈleider also die ˈerste Zeit überˈhaupt ˌnicht*) contributes a second base for this generation. Token 2 is derived from token 1 by removing the central, most strongly palatalized section of the long [nʲnʲ], which shows some energy higher in the spectrum. This reduces the duration of [nʲnʲ] to 140 ms. Token 3 is derived from token 1 by splicing *kann* [kʰan] of g075a000 in place of the original [k̟ʰɛ̈nʲnʲ]. Token 4 is derived from token 3 by also splicing the [a] of [kʰan] into the following *das*, thus replacing the vowel in [əs] *das* by a more peripheral [a]. The splicing disregards the first two and the last period of this vowel to stay clear of the initial aspiration and the final nasalization in [khan]. This reduces the vowel duration marginally by 4 ms from 44 ms in the original *das* to 40 ms.

The f0 contour through [an] of *kann (leider)* (g075a000) is slightly, but perceivably different from the original contour in *kann (Ihnen)*. It has a small dome-shaped rise-fall rather than a continuous

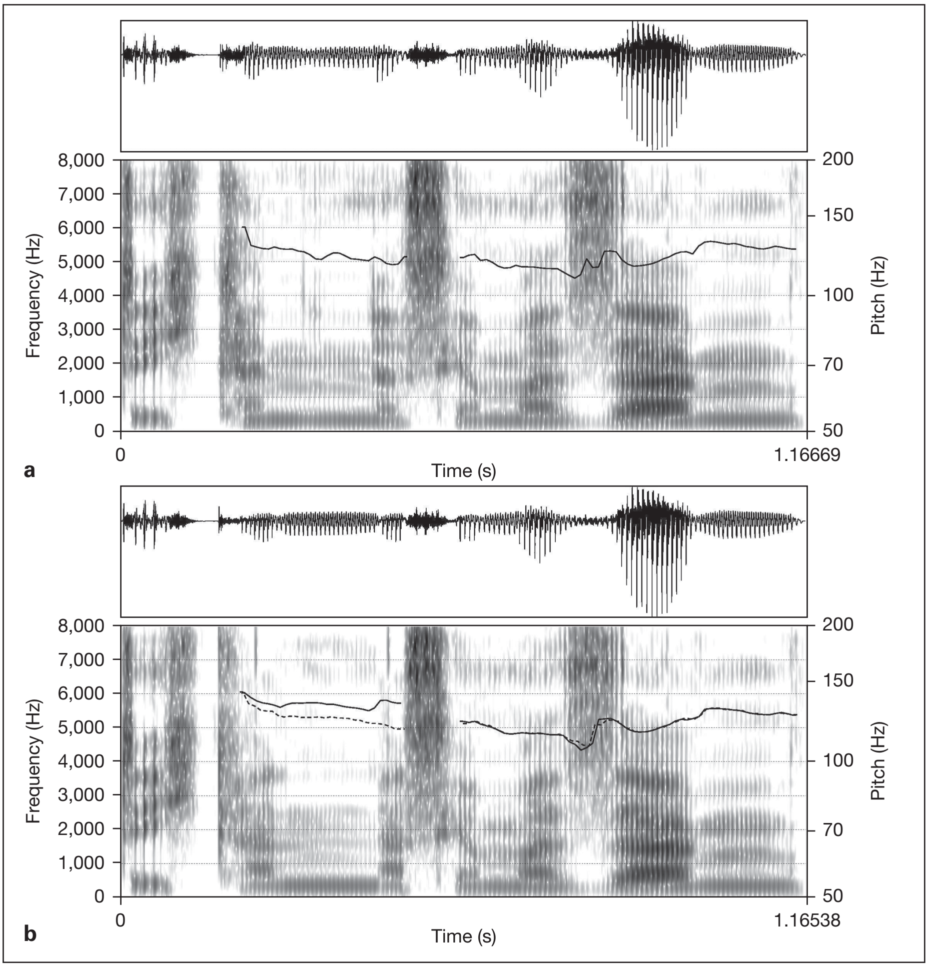

Fig. 6. Speech waves, spectrograms and f0 courses of *ich kann Ihnen das ja mal sagen*, /n/ = 180 ms, fully palatalized (ser1, **a**), fully depalatalized (ser4, **b**). Dotted f0 curve = adjusted f0 (ser6).

fall and is at an overall higher level, although it starts at about the same value (cf. fig. 6, 7). This, together with the longer vowel duration, gives the function word *kann* in the low-pitched sentence frame more prominence and more semantic weight, turning a general possibility into an optional offer by the speaker. This may create a response bias towards accented *kann* and away from *Ihnen.* Auditory examination by the two expert phonetic experimenters indicated that there may be such an effect in the spliced *kann* (*das)* tokens 3 and 4. Therefore, another two tokens were generated from these by transforming f0 point by point to values close to the ones in the original *kann (Ihnen)* (token 5, token 6). This makes f0 comparable across token 1*,* token 2, token 5, token 6, which vary palatality and duration, and it additionally creates a 2 × 2 paradigm of token 5, token 6 versus token 3, token 4 with f0 across *kann* either slightly falling, as in original *kann (Ihnen)*, or slightly rising-falling from *kann (leider)*.

The splicing procedure generates three cornerstones of /n/ durations for subsequent manipulation along a scale from short to long /n/: 180 ms in the original [k̟ʰɛ̈nʲnʲəs] of token 1 (from g072a015),

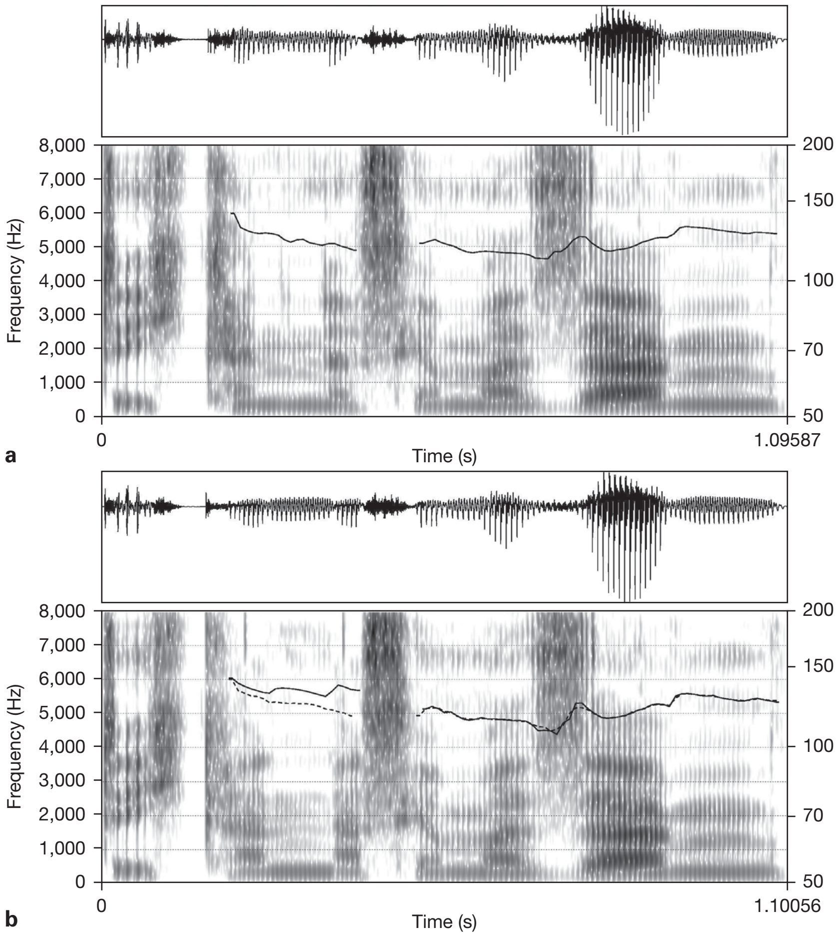

Fig. 7. Speech waves, spectrograms and f0 courses of *ich kann Ihnen das ja mal sagen*, /n/ = 110 ms, fully palatalized (ser1, **a**), fully depalatalized (ser4, **b**). Dotted f0 curve = adjusted f0 (ser6).

140 ms in token 2 from curtailed [k̟ʰĕnʲnʲəs] of g072a015, 110 ms in token 3, token 4*,* token 5, token 6 from [kʰan] of g075a000 + [əs]/[as]. In each of these 6 tokens, 5 durations of /n/ are generated by increasing/decreasing its duration with the Duration Manipulation in *Praat* [Boersma, 2001]: 110, 125, 140, 160, 180 ms. The stretches from dorsal release onset to alveolar fricative onset in these 6 × 5 sections are then spliced at the appropriate zero crossings in the constant frame *ich___s ja mal sagen* of the original utterance g072a015, resulting in the following stimulus series: ser1 – original utterance [k̟ʰĕnʲnʲəs]; ser2 – original utterance minus the central [nʲ] section with higher frequency energy; ser3 – [kʰan] spliced in to replace the palatalized original [k̟ʰĕnʲnʲ]; ser4 – as ser3 but also with [a] for [ə] in *das*; ser5 – f0 modification of ser3 to create a contour comparable to ser1, and ser6 – same f0 modification to ser4 as in ser5.

Figures 6 and 7 compare the fully palatalized (original) and the fully depalatalized (spliced) utterance (ser1, ser4) in their shortest and longest durations of [n], as well as the dome-shaped and falling f0 patterns in the fully depalatalized utterance (ser4, ser6). Depending on the length of [n] and the extent of palatality, the utterances in these series may be decoded as *ich kann das ja mal sagen* or *ich kann Ihnen das ja mal sagen*.

3.3. Test Design

The 6 × 5 = 30 stimuli entered into two different test designs. It is common practice in psycholinguistic listening experiments of this kind to disguise the aspect under scrutiny in the presentation of highly similar stimuli by including 'distractors' that increase variety in the experimental task. Therefore, in test A, 16 distractors were excerpted from the appointment-making files g07 of the same speaker that produced the original test stimulus, 8 with and 8 without *Ihnen* in varying linguistic material, e.g. *ich weiß nicht, wie das morgen (bei Ihnen) aussieht.* 'I don't know how this fits in (with you) tomorrow.' These 16 distractors were copied twice to make the number similar to the test stimuli. The volume was adjusted under auditory control by boosting the signals 6–12 dB in CoolEdit. In all cases, the distractors, like the test stimuli, made sense with or without *Ihnen* so that listeners could be given the task in test A to decide whether the utterances do or do not contain *Ihnen*. All stimuli were copied 4 times and randomized in a test file of 248 utterances, formatted with beeps and 4-second pauses after each stimulus for reactions. When asked for comments on the experiment, several participants pointed out that they felt they had to look for the word *Ihnen* and that it would have been better if the task had been introduced without explicitly referring to it. This indicates that the additional stimuli in the 'present/absent' test paradigm acted as anchors rather than as distractors, because they gave subjects clear cases of stimuli with and without *Ihnen*, against which less clear test stimuli were processed with a bias for the presence of *Ihnen*.

In test B, distractors were not included, and the 6 × 5 = 30 test stimuli were copied 5 times into a randomized and formatted test file of 150 utterances. The actual test was preceded by an audio-visual Powerpoint presentation, which linked the sound of test stimuli that were unequivocally without or with *Ihnen* to the orthographic representations *ich kann das ja mal sagen* or *ich kann Ihnen das ja mal sagen*. If subjects thought that what they heard would be written *ich kann das ja mal sagen*, they were to press button 1 of a computerized reaction time measuring system, if they thought it would be written *ich kann Ihnen das ja mal sagen*, they were to press button 2. This design was to avoid or at least reduce a bias for the presence of *Ihnen*.

3.4. Subjects and Tests

Fourteen subjects did test A, 21 different subjects test B. They were all native speakers of German and students of linguistics. A computerized reaction time measuring system was used that allowed the simultaneous recording of responses and reaction times of up to 8 listeners by pressing one of two buttons on a control box placed on the table in front of each subject. In this system, an impulse triggers a 4-second window to be opened 500 ms into each test stimulus for registering the reaction time up to the point the subject pushes the response button. The earliest position in a stimulus when subjects may be expected to perceive the presence or absence of palatality and therefore the presence or absence of *Ihnen* is after the aspiration of the syllable /ka/, which is 200 ms into each test stimulus. Allowing a minimum of about 300 ms for reaction after perception [a conservative threshold according to Welford, 1980] determines the delay time of 500 ms. If reactions occur within this delay period they cannot have a perceptual grounding and are therefore not recorded. As the test stimuli all measure around 1 s, the 4-second measuring window closes about 3.5 s into the 4-second pause inserted between stimuli in the test file. This time is judged sufficient for a meaningful reaction. If it occurs after the closing of the window the system again ignores it. In both tests, judgements were generally made fairly quickly and in most cases well before the end of the stimulus. This supports the reliability of the experimental data confirming the auditory judgement of both experimenters that the stimuli contained no serious manipulation artifacts.

Subjects did the tests in several small subgroups according to their availability. Each time, pre-recorded instructions on the task and the course of the test were presented from a laptop via loudspeaker. In particular, subjects were asked to keep their right or left hand on the control box all the time to reduce external reaction delay and to respond as quickly as possible. They were also instructed to make a decision even if they were not quite sure. This test introduction included 10 trials to familiarize the listeners with the task and with the equipment. In test B, the instructions were included in the Powerpoint presentation. The tests took place in the sound-treated studio of the IPDS Kiel, and the stimuli were presented through loudspeaker. The test section in test A lasted 30 min, in test B 18 min, plus approximately 10 min for instructions in each case.

4. Results and Statistics

The results of both tests were analyzed in terms of descriptive and inferential statistics. The latter included separate repeated-measures ANOVAs for the judgement and the reaction-time data. The two ANOVAs were based on the fixed factors ‘series’ (6 levels) and ‘nasal duration’ (5 levels). In the case of the judgement data the dependent variable was the *Ihnen* frequency obtained for each subject and stimulus across all 5 repetitions. By summing responses, the binary *Ihnen* judgements became metrical values between 0 and 5. In a structurally analogous procedure, the reaction-time data were converted into single measurements for each subject and stimulus by averaging the values of the 5 repetitions. The resulting means then served as the dependent variable. The sample sizes of the two ANOVAs were n = 14 in test A and n = 21 in test B. In each ANOVA, the two fixed factors, as well as the interactions between them, violated the sphericity criterion (cf. Mauchly test). Therefore significance values reported below were based on Greenhouse-Geisser corrections.

4.1. Results of Test A

All 14 subjects responded to the test stimuli within the given reaction time window of 4 s: only for 3 out of the total of 1,680 was there no response record. Disregarding this extremely low percentage, it was assumed that judgements were reliable and suitable for statistical analysis. Figures 8 and 9 show the descriptive results of *Ihnen* judgements and reaction times in test A. Each bar is based on 56 responses, or exceptionally on 55 in the rare case of a missing judgement. The organization of figures 8 and 9 in 6 groups of 5 bars parallels the make-up of the ANOVAs.

The overall picture of figure 8 can be summarized in terms of four characteristics: (1) The stimuli of ser1, which contained the original [n^jn^j], as well as the stimuli of ser2, in which the central part of increased palatalization was excised from the original [n^jn^j], were judged in almost 100% of the cases as containing *Ihnen*. (2) The remaining ser 3–6, where palatality was reduced beyond the nasal into *ka*(*nn*) and (*d*)*a*(*s*), differ from ser1–2, with *Ihnen* judgements only around 40–60%. (3) ser2–6 show an increase in *Ihnen* judgements with lengthening of the nasal portion. The *Ihnen* judgements differ by up to 30% between the longest and the shortest nasal portions. (4) There are no obvious differences between the judgement patterns of ser3 and ser5 on the one hand and ser4 and ser6 on the other, despite the different f0 courses on *kann*.

The two main effects ‘series’ and ‘nasal duration’ are highly significant in the judgement data with similar effect sizes (partial eta squared, η^2_p): series [$F(5, 65) = 14.826$; $p = 0.001$; $\eta^2_p = 0.533$]; nasal duration [$F(4, 52) = 20.506$; $p < 0.001$; $\eta^2_p = 0.612$]. Among the pairwise comparisons of levels of the factor ‘series’ (with

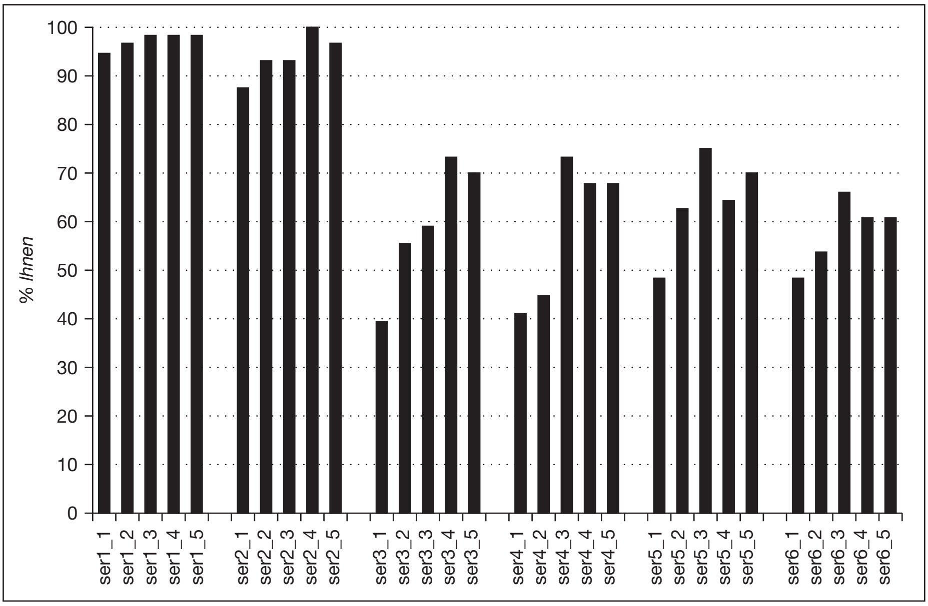

Fig. 8. Percentages of *Ihnen* judgements obtained for the 6 × 5 experimental stimuli of test A across all 14 subjects and 4 repetitions, i.e. n = 56 (55) for each bar.

Bonferroni corrections included in the significances), ser1 and ser2 are significantly different from all other series at $p < 0.05$, but there are no significant differences between ser1 and ser2 nor among ser3–6. In similar pairwise comparisons, duration level 1 differs from levels 3–5 at $p < 0.01$, duration level 2 from levels 3–5 at $p < 0.05$; no others are significant.

These results show that ser1–2 and ser3–6 with strong and weak palatality, respectively, form two perceptual groups. The descriptive data of figure 8 indicate that the influence of nasal duration is prominently linked to ser3–6, where levels 1 and 2 differ from the remaining ones. This tie of nasal duration with series is also mirrored by the significant interaction of the two factors, albeit with a much smaller effect size: F (20, 260) = 2.341; p = 0.038; $\eta^2_p = 0.153$.

Figure 9 provides an overview of the means and standard deviations of the reaction times to the individual stimuli across the 14 subjects. The repeated-measures ANOVA yields a highly significant main effect of series [$F(5, 65) = 35.789$; $p < 0.001$; $\eta^2_p = 0.734$], but no significant effect of nasal duration [$F(4, 52) = 1.803$; $p = 0.202$; $\eta^2_p = 0.097$], and only a marginally significant interaction between the two fixed factors [$F(20, 260) = 6.927$; $p = 0.021$; $\eta^2_p = 0.348$]. As for pairwise comparisons, ser1 is significantly different from ser2 at $p = 0.05$, and from all other series at $p < 0.01$, likewise ser2 at $p < 0.05$. There are no significant differences among ser3–6. So, the reaction-time profiles across the 6 series reflect the response profiles, but although the comparison of the response data of ser1–2 is not significant, the corresponding comparison of the reaction times is marginally so. This suggests that the highest degree of palatality is mirrored in the shortest reaction times. There are no observable regularities

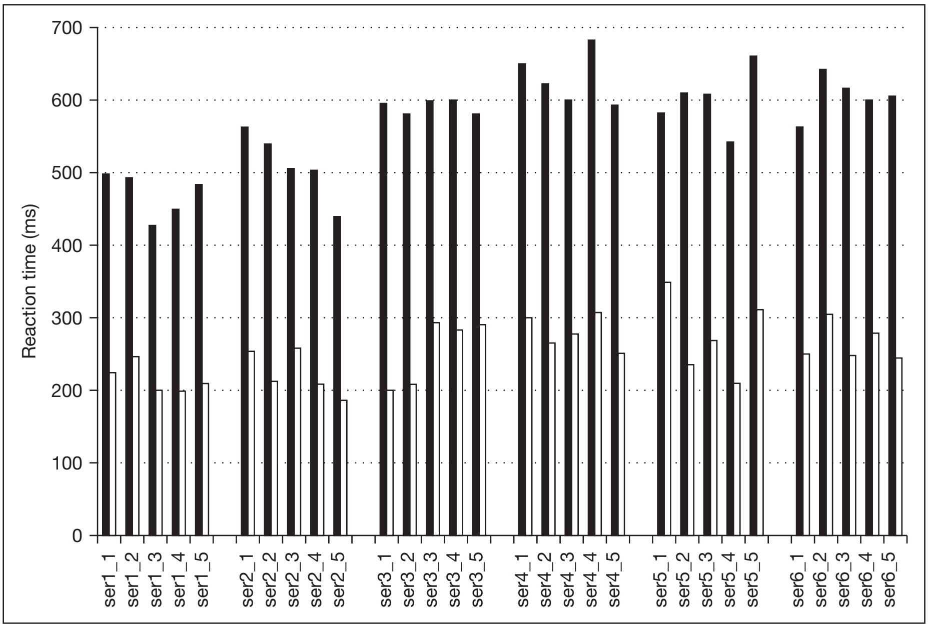

Fig. 9. Reaction times (in ms) of the 6 × 5 experimental stimuli in test A. Black and white bars represent means and standard deviations, respectively. Each bar is based on 56 (55) reactions.

in the patterning of the duration levels in figure 9, which is also supported by none of the pairwise comparisons reaching significance.

4.2. Results of Test B

It is obvious from figure 8 that the majority of stimuli in test A were judged as containing *Ihnen* in more than 50% of cases. This holds even for those stimuli in which all relevant, palatalized sound sections were removed or replaced and in which the two experimenters themselves were unable to perceive the word *Ihnen*. Obviously, the test paradigm made subjects want to find the word in the stimuli. Such hypersensitivity towards *Ihnen* may mask weaker prominence-related effects. Therefore test B deviated from the established psycholinguistic design by not including distractor stimuli, and the subjects' task was to relate the stimuli to orthographic representations of sentences, hence avoiding an explicit verbal reference to the presence/absence of the target word *Ihnen*.

The descriptive analyses of the judgement and reaction-time data are summarized in figures 10 and 11. Each bar conflates the data of 21 subjects and 5 repetitions and hence represents 105 values or, in exceptional cases, 104 values due to missing responses. However, as in test A, the lack of responses was negligible: only 4 out of 3,150 judgements. The organization of figures 10 and 11 in 6 groups of 5 bars parallels the make-up of the ANOVAs.

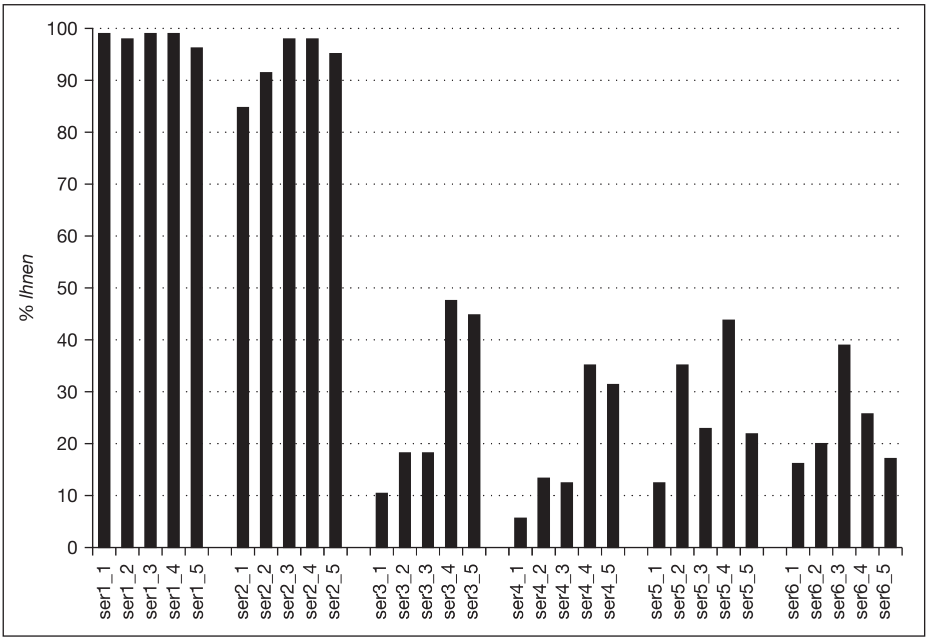

Fig. 10. Percentages of *Ihnen* judgements obtained for the 6 × 5 experimental stimuli of test B across all 21 subjects and 5 repetitions, i.e. n = 105 (104) for each bar.

4.2.1. Identification Judgements – Overview of all Stimulus Series

The two main factors 'series' and 'nasal duration' come out clearly significant for the judgement data. However, in terms of effect size (partial eta squared, η^2_p), the duration of the nasal [$F(4, 80) = 10.660$; $p < 0.001$; $\eta^2_p = 0.509$] is only about half as influential as the stimulus series [$F(5, 100) = 77.490$; $p < 0.001$; $\eta^2_p = 0.920$]. Figure 10 shows that the effect of the stimulus series is mainly due to the difference between ser1–2 on the one hand and ser3–6 on the other. While *Ihnen* perception dominates in the former two series, it diminishes considerably in the latter four series. This bipartition is supported by general post-hoc tests that were performed (with Bonferroni corrections included in the significances) between all levels of the factor 'series'. ser1 and ser2 do not differ significantly from each other, but do differ from all other series (with $p < 0.001$). In this respect, the results of test B are congruent with those of test A.

However, the stimuli of ser3–6 triggered considerably fewer *Ihnen* judgements in test B than in test A. The *Ihnen* percentages of ser3–6 are consistently well below 50% (even as low as 5–16% in the majority of stimuli) in test B, and hence the difference between ser1–2 and ser3–6 is sharper in test B. In terms of maximum differences of *Ihnen* percentages across all stimulus series, test B yields a value of 94.2%, compared with 60.8% in test A.

With regard to the effect of nasal duration, there is an overall increase in *Ihnen* perception for higher stimulus numbers, i.e. for longer nasal sections, as in test A. But unlike in test A, there is also a consistent decrease from stimulus 4 to 5. The bidirectional

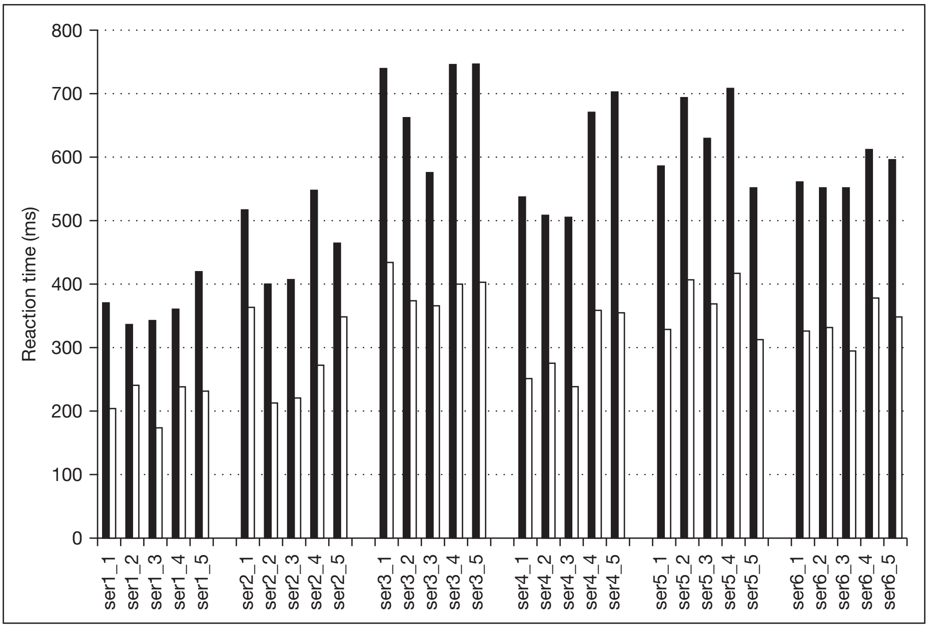

Fig. 11. Reaction times (in ms) of the 6 × 5 experimental stimuli in test B. Black and white bars represent means and standard deviations, respectively. Each bar is based on 105 (104) reactions.

Ihnen changes across the stimulus series are also manifest in the general post-hoc tests that were performed between the levels of the factor 'nasal duration'. Stimuli 1 and 4 differ from each other ($p < 0.01$) and from all other stimuli, yielding, on average, the lowest and highest numbers of *Ihnen* judgements, respectively. But no significant differences result for the comparisons of the two centre stimuli 2 and 3 with each other and with stimulus 5.

Figure 10 also shows that there are two exceptions to the predominant result pattern of the factor 'nasal duration'. First, ser1 does not show an increase, only a small decrease of *Ihnen* judgements. Second, the *Ihnen* decrease is stronger for ser5–6. In ser6, the decrease already applies to stimulus 4 and is almost as strong as the preceding increase. These exceptions are reflected in a significant interaction between the two main effects of series and nasal duration [$F(20, 400) = 2.682$; $p < 0.001$; $\eta^2_p = 0.261$].

4.2.2. Identification Judgements – Detailed Analysis of Stimulus Series 3–6

Comparing figure 10 with figure 8 shows that *Ihnen* identification in ser3–6 of test B differs from test A. In test B, there is a significant difference in the comparison of ser3 and ser4 ($p = 0.046$). The difference is due to overall fewer *Ihnen* judgements in ser4, corresponding to the further decrease of palatality from ser3 to ser4. As can be seen in figure 10, the same relationship holds for ser5 and ser6, but the lowered amount of *Ihnen* identifications in ser6 is not clear enough for the effect to become significant.

Furthermore, there is an increase for stimuli 1–3 from ser3 to ser5 and likewise from ser4 to ser6, but a decrease for stimuli 4–5 from ser3 to ser5 and likewise from ser4 to ser6. However, the post-hoc comparisons of the ANOVA do not take the possibility into account that further *Ihnen* differences may exist between *subsets* of stimuli because within each of the fixed factors 'series' and 'nasal duration' the judgement data are pooled across all levels of the respective other factor, potentially masking significant differences between levels. Moreover, pairwise comparisons of levels *across* the two factors are not included, but such crosswise comparisons of factor levels are necessary to analyze the results comprehensively with regard to the prominence-related hypothesis 3.

For this reason, a supplementary statistical test was applied to the data of test B. It examined the effects of nasal duration separately for the two stimulus groups 1–3 and 4–5, based on cross-series comparisons within ser3–6. Due to the smaller amount of data, the conservative non-parametric Wilcoxon-Wilcox multiple comparisons test was used [Sachs, 1972]. The organization of the stimulus groups takes into account that stimuli 1–3 contain shortish nasals that are separated from each other by just 15 ms, whereas the nasal durations of stimuli 4–5 increase in steps of 20 ms and are a good deal longer overall. For each stimulus group and series level the *Ihnen* frequencies were summed across all 21 subjects. These sums, which varied between 0 and 315 for stimulus groups 1–3 and between 0 and 210 for stimulus groups 4–5, provided the basis for the multiple comparisons. The sample size was $n = 21$ in all comparisons. The number of compared conditions was $k = 4$.

The results of the Wilcoxon-Wilcox test show in addition to the post-hoc tests of the ANOVA that stimuli 1–3 of ser4 together yield significantly fewer *Ihnen* responses than stimuli 1–3 of ser6 ($RD = 27 > 26$; $p = 0.01$, two-tailed) and of ser 5 ($RD = 22.5 > 21.5$; $p = 0.05$, two-tailed); the lower number of the *Ihnen* responses for stimuli 1–3 from ser3 compared to ser5 can count as marginally significant ($RD = 21 < 21.5$; $p = 0.05$, two-tailed). As regards the groups of stimuli 4–5, there are more *Ihnen* responses in ser3 than in ser5 ($RD = 22 > 21.5$; $p < 0.05$). Additionally, the number of *Ihnen* responses of stimuli 4–5 is significantly higher in ser6 than ser4 ($D = 24 > 21.5$; $p = 0.05$, two-tailed) and from ser3 to ser6 ($RD = 40 > 26$; $p = 0.01$, two-tailed).

Combining the non-parametric cross-factor comparisons with the within-factor post-hoc comparisons of the ANOVA reveals for ser3–6 that (1) the *more palatalized* ser3 yields *more Ihnen* responses than the less palatalized ser4, that (2) the *shorter* stimuli 1–3 trigger *fewer Ihnen* responses in ser3 and ser5 than in ser4 and ser6, and that (3) the *longer* stimuli 4–5 trigger *more Ihnen* responses in ser3 and ser5 than in ser4 and ser6.

4.2.3. Reaction Times

Figure 11 provides an overview of the means and standard deviations of the reaction times to the individual stimuli across the 21 subjects. The corresponding repeated-measures ANOVA yields highly significant main effects of both series [$F(5, 100) = 44.681$; $p < 0.001$; $\eta^2_p = 0.691$] and nasal duration [$F(4, 80) = 15.428$; $p < 0.001$; $\eta^2_p = 0.435$], as well as a significant interaction between the two fixed factors [$F(20, 400) = 6.076$; $p < 0.001$; $\eta^2_p = 0.233$]. Figure 11 shows that the significant main effect of the fixed factor 'series' is linked to a clear progression in reaction times from ser1 to ser2 to ser3–6. While the stimuli of ser3 require the longest reaction times, i.e. 650–750 ms on average, the reaction times of ser1 stimuli are only about half as long. So, there is a correspondence between the *Ihnen* judgements and the reaction times. The stimuli that trigger almost exclusively *Ihnen* identifications (ser1–2) have shorter reaction times than the stimuli that are predominantly perceived without *Ihnen* (ser3–6); The general post-hoc tests (with Bonferroni corrections) return significant differences for ser1 and ser2 between each other and with all other series (with $p < 0.001$). In this respect, test B coincides with test A, but as in the response data, the break between the groups ser1–2 and ser3–6 is sharper in test B.

Within ser3–6, stimuli with more ambiguous *Ihnen* identifications (closer to 50%) are judged more slowly than stimuli with little or no *Ihnen* identifications: ser3,5 versus ser4,6.

Reaction-time differences are significant in the following pairwise comparisons: ser3 versus ser4 ($p = 0.001$), ser3 versus ser6 ($p = 0.001$), ser5 versus ser6 ($p = 0.022$). The comparison between ser4 and ser5 shows a trend towards significance ($p = 0.091$). The overall significance pattern within the factor 'series' matches well with the stepwise increase in reaction times from ser1 to ser2 and further from ser4,6 to ser3,5 in the descriptive account. Thus the reaction time data of test B differ from those of test A in the presence of significantly different reaction-time levels among ser3–6.

As regards the effect of the fixed factor 'nasal duration', the reaction time means in figure 11 tend to be U-shaped across stimuli 1–5, except in ser5, which reflects the significant interaction of 'nasal duration' with 'series'. These profiles indicate that on average the stimuli at either end of the nasal duration scale cause higher reaction times than the stimuli in the centre, where reaction times are up to 20% (or about 200 ms) lower. The U-shapes are also mirrored in the post-hoc pairwise comparison tests. For the factor 'nasal duration' significantly different reaction-time levels are found between stimuli 2 and 4, 2 and 5, 3 and 4, and 3 and 5 ($p < 0.01$). The difference between the higher reaction time of stimulus 1 and the lower reaction time of stimulus 3 shows a trend towards significance ($p = 0.082$). In contrast, there are no significant reaction time differences between peripheral stimuli of the series, i.e. between 1 and 4, 1 and 5, and 4 and 5. Test B differs from test A as to the significance of the factor 'nasal duration' and the presence of a U-shaped profile of the five duration levels within each series (except ser5).

5. Interpretation of the Results and Evaluation against the Hypotheses

5.1. Identification Judgements

The descriptive and inferential statistics of the perception data show very clearly that the decoding of the utterances as either containing or not containing *Ihnen* depends on an articulatory prosody of palatality across *kann__das*. The break in the perceptual profile occurs between ser1–2 on the one hand and ser3–6 on the other, due to the successive removal of palatalization in the nasal, of fronting in the plosive friction, and of the raised and central quality of the vowel in /kan/. These findings apply to the results of both test A and test B, although the break is sharper in test B than in test A. An articulatory prosody of palatality is thus a robust cue in perceptual identification, irrespective of the experimental design. However, the descriptive data of test A allow no further differentiation of the extent of the articulatory prosody of palatality in ser3–6, whereas the data of test B show much finer gradation cued by the further replacement of the central vowel by a lower and more peripheral one in *das* of ser4,6 versus ser3,5.

The data of both tests also show the influence of nasal duration on *Ihnen* identification, more prominently so for ser3–6 than for ser1–2. Its perceptual strength is, however, much lower, as is evidenced by the lower partial eta squared for the duration than for the series effect in test B, and by the very low partial eta squared for the interaction between the two effects in both tests.

In test B, unlike in test A, the significant effects point to a partition of the duration scale into three domains with regard to *Ihnen* identification, i.e. (1) stimulus 1, (2) stimuli 2, 3, 5, and (3) stimulus 4. Whereas the decrease of the articulatory prosody of palatality from ser1 to ser4 results in a successive decrease of *Ihnen* judgements, the factor of nasal duration has the effect of raising *Ihnen* judgements across the stimuli, but lowering them again to stimulus 5. The descriptive data show very little influence of nasal duration for ser1 and ser2, whilst the decrease of *Ihnen* judgements for the high stimulus numbers is most pronounced in ser5–6. These facts point to the interaction of the factors 'series' and 'duration', with 'series' dominating 'duration'. When palatality is strong (ser1,2), nasal duration has very little influence on *Ihnen* judgements; when palatality is weak (ser3,5) or absent (ser4,6), duration can only weakly compensate for it.

The significant decrease of *Ihnen* responses from stimulus 4 to stimulus 5 in all series of test B may be linked to increased prominence in *kann.* Auditory examination, by the two authors, of stimuli 4 and 5 in each series suggests that stimulus 5 gives more prominence to *kann*, due to extreme nasal duration, thus triggering a deflection of [n] duration from the decoding of *Ihnen* to a new semantic weighting of *kann.* This lowers the frequency of *Ihnen* responses.

The finer effects in the nasal duration groups of ser3–6 add a more detailed picture to the results of test B. In ser3 and ser4 as against ser5 and ser6, the *shorter* nasal durations of stimuli 1–3 trigger *fewer* responses of *Ihnen*, the *longer* nasal durations of stimuli 4–5 trigger *more*. The phonetic explanation can be sought in the stimulus generation of ser3–6. The base stimulus in ser3 and ser4 has a nasal duration of 110 ms, which is stretched to 160 and 180 ms in stimuli 4 and 5, turning the dome-shaped f0 pattern in the nasal into a highish plateau (compare fig. 6 with fig. 7). The resulting expansion of high pitch ties in with the microprosodic raising of f0 in high vowels and in palatalized sonorants, and with high-frequency spectral energy in high vowels, and thus leads to an increase in palatality. As a result, the number of *Ihnen* judgements goes up, compared with the corresponding stimuli 4, 5 in the equivalent palatality of ser5 or ser6, respectively, where the stretching of a falling f0 pattern in the nasal of the base stimulus does not create the high pitch environment. In the shorter nasal durations of 110–140 ms, the dome-shaped f0 pattern is not expanded to such a long highish f0 plateau. But the dome-shape gives the word *kann* clearly perceivable prominence in the utterance frame, which is absent when the word is combined with the falling f0 pattern. Since prominence is also connected to lengthening, the perceptual processing of the duration of the nasal will be reorganized from linking it with the lexical item *Ihnen* to associating it primarily with prominence of *kann*. The result is a decrease of *Ihnen* judgements in the equivalent palatality of ser3 versus ser5, and of ser4 versus ser6.

In this phonetic interpretation, dome-shaped f0 increases the prominence of *kann* for the shorter nasal durations, which in turn reduces the cue value for *Ihnen*. On the other hand, it strengthens the articulatory prosody of palatality by high pitch in the long nasal durations, thus weakening the general prominence effect of extreme nasal lengthening in stimulus 5 of all series: the decrease of *Ihnen* responses from stimulus 4 to stimulus 5 is smaller in the dome-shaped f0 of ser3,4 than in the falling f0 of ser5,6. This line of argument is further supported by the following comparisons. The short nasal stimuli of ser4 differ from those of ser5 by less palatality *and* by dome-shaped f0, both decreasing *Ihnen* responses in the short duration group (stimuli 1–3), resulting in a significant effect between the two series. Similarly, the long-nasal stimuli of ser3 differ from those of ser6 by more palatality *and* by dome-shaped f0, both increasing *Ihnen* responses in the long duration group (stimuli 4–5), resulting in a significant effect between the two series. In these cases, the interactive effects of degree of palatality and f0 for *Ihnen* responses are additive. On the other hand, in ser3 versus ser4 and ser5 versus ser6 stimuli have the same f0 pattern, but differ in the degree of palatality. So, the differential influence of f0 on the two duration groups cannot apply; the degree of palatality affects the entire duration scale in the series, and the two duration groups do not show significance for the two series pairings.

In summary, we can say that both in test A and test B the articulatory prosody of palatality has the strongest effect on *Ihnen* judgements. When palatality extends across the whole of *kann*__*das*, irrespective of the presence or absence of the most palatalized nasal section, the duration of the nasal becomes negligible. If *kann*__*das* is depalatalized, nasal duration shows a weak effect that cannot compensate for reduced palatality, even with the longest duration. In test B the effect of nasal duration is influenced by additional effects of prominence at the upper end of the nasal duration scale in all series and in the lower part of the scale in ser3–6, as well as by a microprosodic effect in the upper part of the sale in ser3–6. The former prosodic effect decreases *Ihnen* judgements, the latter increases them.

The differences between the data of test A and test B now raise four questions: (1) Why is the proportion of *Ihnen* judgements higher across all series in test A? (2) Why is the divide between ser1–2 and ser3–6 less sharp in test A? (3) Why is the finer differentiation in ser3–6 absent from test A? (4) Why is there no significant difference between stimuli 4 and 5 across all series?

In each case, the answer can be found in the different test paradigms. Test A followed established psycholinguistic procedure by including 'distractors' which had to be fitted into the design of the listening experiment that required subjects to press one of two buttons for presence or absence of *Ihnen*. Thus the distractors became cornerstone stimuli, providing clear cases for one or the other response in a balanced set between these anchors and the manipulated cases. In turn, the task had to be formulated as a response to '*Ihnen* present/not present'. This made listeners keen to hear *Ihnen*. Consequently, the number of false alarms went up across all stimulus series, weakened the divide between the two groups of strong and weak palatality and blurred fine differences in the weak-palatality group and between stimuli 4 and 5 in all series.

5.2. Reaction Times

The mean reaction times of the *Ihnen* judgements in test B show a significant stepwise increase from ser1 to ser2 to ser3–6. The largest step occurs between ser1–2 on the one hand and ser3–6 on the other and hence coincides with the major drop in *Ihnen* percentages.

The reaction-time step from ser1 to ser2 is not paralleled by a significant change in *Ihnen* judgements. Yet, the increasing reaction times indicate that the stimuli of ser2, in which the nasal section with the strongest palatalization was cut out, were more ambiguous in signaling *Ihnen* than the derivatives from the naturally produced stimulus in ser1. These patternings also apply to the data of test A, but the divide between ser1–2 and ser3–6 is again less sharp, matching the difference in the judgement data between the two tests.

Within ser3–6 of test B, mean reaction times decrease again in two steps from ser3 to ser4 and from ser5 to ser6. As the f0 patterns do not differ within ser3–4 and ser5–6, the changes in reaction times can be ascribed to the different residuals of palatality in these pairs of series. While the palatality in [kann] was removed in all four series, ser3,5 still have the central vowel in *das*, but there is a more open and more peripheral vowel in ser4,6, moving still further away from the palatality of *Ihnen*. Hence, as regards *Ihnen* perception, there is a cue conflict between [kann] and the following [ə] in ser3, which is resolved in [kannas] in ser4. The resulting more consistent cues go along with lower reaction times. The same holds for ser5 and ser6. However, as in the case of ser1 and ser2, the reaction time difference between ser5 and ser6 is not mirrored in a significant change of *Ihnen* judgements. These finer patternings of reaction times in ser3–6, like the corresponding response patterns, are absent from test A, due to the experimental design.

Within a series of test B, mean reaction times are significantly lower for the stimuli with moderate nasal durations than for the stimuli with extreme nasal durations. The effect can be found in all series, including the fully depalatalized ones. The decoding of the nasal durations at the upper end of the duration scale has been related to the additional effect of prominence, due to extreme lengthening, which triggers a deflection of [n] duration from the decoding of *Ihnen* to a new semantic weighting of *kann* (cf. 5.1). This lowers the frequency of *Ihnen* responses from stimulus 4 to stimulus 5 in all series, and the parallel decoding keeps the reaction time for both stimuli across all series at a similar high level. The series-internal reaction time patterning is not found in test A, due to the experimental design.

5.3. Hypotheses

Hypothesis 1 (effect of palatality) has been confirmed by the results of both test A and test B in that decoding the utterance as containing *Ihnen* decreases with the successive reduction of palatality from its extension across the whole stretch of [k̟ʰɛ̈nʲnʲəs] in the original, except for the lack of an effect in judgement data when only the most palatalized part of the nasal is removed. However, the significant increase of reaction times indicates that this reduction of palatalization already introduces more uncertainty.

Hypothesis 2 (effect of nasal duration) has not been confirmed in its general assertion that decoding the utterance as containing *Ihnen* can be assumed to decrease with the shortening of the nasal consonant duration within the different frames of palatality. In both tests, the effect is series-dependent, not monotonic, and generally weak. In test B, it is influenced by additional effects of prominence at the upper end of the nasal duration scale in all series and in the lower part of the scale in ser3–6, as well as by a microprosodic effect in the upper part of the scale in ser3–6. However, it needs to be recognized that what has been called 'short' and 'long' is relative to the manipulated duration scale from 110 to 180 ms. In absolute terms, even 110 ms is a long duration in intervocalic position of a sequence of two unaccented function words. In this environment, even 75 ms in *kann das* relates to geminate [kʰannəs] (cf. 3.1). The 58 ms of the intervocalic nasal in *kann ich* (cf. 3.1) is a representative duration for an intervocalic singleton [n]. Therefore the following hypothesis may be advanced for further testing: if duration is manipulated in the range between 50 and 120 ms in the different palatality series, duration will show a strong effect in a three-way differentiation between *kann Ihnen das/kann das/kann es ja mal sagen.*

Hypothesis 3 (effect of prominence) has been partly confirmed in test B. It has been established that (1) a dome-shaped as against (2) a continuously falling f0 contour in [kʰanna/əs], combined with the two lowest degrees of palatality, generates different degrees of prominence on *kann* and results in different response profiles for *Ihnen* with the shorter nasal durations. However, for dome-shaped f0 with the corresponding long nasal durations, a prominence effect has not surfaced in the data analysis. The expanded high f0 pattern at long nasal durations rather leads to a microprosodic effect of increased high pitch, strengthening the effect of palatality, and resulting in an increase of *Ihnen* judgements from (1) to (2). On the other hand, an additional prominence effect emerged, related to the longest nasal duration at all degrees of palatality.

6. Conclusion and Outlook

The central aim of this paper has been the investigation into the articulatory prosody of palatality as a factor in the perception of reduced function words in German, thus complementing the existing accounts of function word production, within an overall theoretical framework of phonetic reduction in speech communication. It leads on from the analysis of this articulatory prosody in another German function word in Niebuhr and Kohler [in press]. Its results show again that an articulatory prosody is a highly significant cue to the decoding of utterances that are lexically differentiated. Phonetic detail is thus essential in the perception and the cognitive processing of speech.

The function word *Ihnen* can be realized in a number of different ways from elaborated to highly reduced, depending on situational and phonetic environments. The weakly reduced form [iːnʲnʲ] and the more strongly reduced form [nʲnʲ] 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 nasality, as do other intermediate degrees of reduction. This means that all phonetic forms of this word can be conceptualized as containing these features; they constitute the ‘phonetic essence’ [Niebuhr and Kohler, in press] of *Ihnen*. This concept of phonetic essence may be assumed to apply to function words generally. Such a phonetic essence of a lexical item manifests itself as a gestalt feature of opening-closing gestures, either in segmentable 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, [k̟ʰɛ̈nʲnʲəs], as against [kʰannas] *kann das*, can be decoded as containing *Ihnen* although there is no semantic bias in the same linguistic and situational context.

In their investigation into the perceptual relevance of a prosody of palatality in extreme reduction of the German particle *eigentlich*, Niebuhr and Kohler [in press] argued that all the realizations of *eigentlich* form a class that is characterized by the phonetic essence of nasality and of palatality and their spread across the word from the gliding portion of the word-initial diphthong. In the most extreme reduction of the word, [aĩ̃], this palatality is compressed into the palatal gliding of the diphthong. The palatal gliding duration then becomes a distinctive differentiator between [aĩ̃nʲə] *eigentlich 'ne* 'actually a' and [aĩnə] *eine* 'one'. The manipulation of this duration proved highly relevant for the perception of the word.

In the findings of the present investigation, palatality is again spread across several syllables. But here the extension of palatality was manipulated two-dimensionally, by exchanging palatal and non-palatal syllables and by stepwise duration changes in differently palatalized nasal consonants, compared with the one-dimensional manipulation of palatality in the diphthongal gliding of *eigentlich*. The duration of the palatalized nasal had little effect within the duration range that was manipulated, as long as the other palatal syllable features were kept. However, as has been pointed out, the duration scale used for stimulus generation needs to be extended at the lower end in a future experiment. The hypothesis is that the relationship between palatality and its duration in the phonetic essence of *Ihnen* will then turn out to be comparable to the one in the phonetic essence of *eigentlich*; in both cases the graded duration of palatality, a palatalized nasal in one, a palatal glide in the other, are expected to be important for lexical decoding over and above the mere presence or absence of palatality.

To introduce such phonetic detail in experimental designs it is necessary to take naturally produced utterances as a point of departure, preferably, as was done for this paper, spontaneous utterances that were recorded for purposes outside the particular research question. This methodology allowed the generation of stimuli in which it was possible to vary the extension of palatality by splicing auditorily assessed parts of utterance from the same speaker produced in a spontaneous speech corpus. Special care was taken to guarantee natural sounding stimuli without acoustic artifacts as a valid basis for perception experiments. The application of this methodology has been successful. The experiments have also highlighted another methodological aspect. The use of distractors and the wording of the task to be performed by the listeners need to be considered very carefully and to be adapted to the particular research question and test design to avoid response bias and the masking of fine perceptual effects. The strategy adopted in test B has successfully dealt with these issues.

A substantial percentage of responses were made early in the reaction window, i.e. within a short reaction delay after /ka/. This suggests that listeners took a decision on the strength of having received either the palatalized or the non-palatalized syllable /ka/ and before the nasal residue of *Ihnen*. Moreover, an increase of *Ihnen* responses also occurred when vowel raising and centralization as an indicator of palatality was restricted to the vowel /a/ after /kann/, and thus outside the segmental residue of *Ihnen*. These facts further underscore the importance of non-segmental articulatory prosodies and argue against the conception of speech perception on the basis of a linear phonemic string. The response profiles also point to an influence of pitch patterns on the processing of lexical information in utterances. The results furthermore confirm that articulatory prosodies can trigger lexical information directly, i.e. without help from syntactic and semantic contextualization. In view of these findings, future research needs to take a critical look at the traditional segment-prosody dichotomy in two ways. Since articulatory components of segments assume prosodic extension in articulatory prosodies, they need to be focussed on, and since segmental categorization is embedded in prosodic patterns (in the traditional meaning of prosody), the divide between the two levels has to be made more permeable.

Other articulatory prosodies that have been identified in the study of connected speech production in communicative interaction, e.g., velarization, nasalization, lip rounding, also await being investigated as to their influence in the perception and cognitive processing of speech, and the prosodic research paradigm will also have to include a broader spectrum of the lexicon and will have to be applied to a variety of languages.

Acknowledgement

The authors are very grateful to Ernst Dombrowski from the Department of Psychology of Kiel University for his helpful advice on statistical methodology and analysis, and to Bill Barry as well as to two reviewers for valuable suggestions for revision.

References

Boersma, P.: Praat, a system for doing phonetics by computer. Glot int. *5:* 341–345 (2001).

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

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

Byrd, D.; Kaun, A.; Narayanan, S.; Saltzman, E.: Phrasal signatures in articulation; in Broe, Pierrehumbert, Papers in Laboratory Phonology V, pp. 60–87 (Cambridge University Press, Cambridge 2000).

Ernestus, M.; Baayen, H.R.; Schreuder, R.: The recognition of reduced word forms. Brain Lang. *81:* 162–173 (2002).

Firth, J.R.: Sounds and prosodies. Trans. Philological Soc. *1948:* 127–152 (1948).

Hawkins, S.: Roles and representations of systematic fine phonetic detail in speech understanding. J. Phonet. *31:* 373–405 (2003).

Hawkins, S.; Nguyen, N.: Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. J. Phonet. *32:* 199–231 (2004).

Hawkins, S.; Smith, R.: Polysp: a polysystemic, phonetically-rich approach to speech understanding. Ital. J. Ling. *13:* 99–188 (2001).

IPA: Handbook of the International Phonetic Association (Cambridge University Press, Cambridge 1999).

IPDS: The Kiel corpus of read speech, vol. I, CD-ROM No. 1 (IPDS, Kiel 1994).

IPDS: The Kiel corpus of spontaneous speech, vol. I, CD-ROM No. 2 (IPDS, Kiel 1995).

IPDS: The Kiel corpus of spontaneous speech, vol. II, CD-ROM No. 3 (IPDS, Kiel 1996).

IPDS: The Kiel corpus of spontaneous speech, vol. III, CD-ROM No. 4 (IPDS, Kiel 1997).

Jones, D.: An outline of English phonetics; 8th ed. (Heffer & Sons, Cambridge 1956). (First published 1918).

Kelly, J.; Local, J.: Doing phonology (Manchester University Press, Manchester 1989).

Kemps, R.; Ernestus, M.; Schreuder, R.; Baayen, H.R.: Processing reduced word forms: the suffix restoration effect. Brain Lang. *90:* 117–127 (2004).

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

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

Kohler, K.J.: Kommunikative Aspekte satzphonetischer Prozesse im Deutschen; in Vater, Phonologische Probleme des Deutschen, pp. 13–39 (Narr, Tübingen, 1979).

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

Kohler, K.J.: Segmental reduction in connected speech in German: phonological facts and phonetic explanations; in Hardcastle, Marchal, Speech production and speech modelling, pp. 69–92 (Kluwer Academic Publishers, Dordrecht 1990).

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

Kohler, K.J.: Gestural reorganization in connected speech: a functional viewpoint on ‘Articulatory Phonology’. Phonetica *49:* 205–211 (1992).

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

Kohler, K.J.: The disappearance of words in connected speech. ZAS Working Papers Ling. *11:* 21–34 (1998).

Kohler, K.J.: Articulatory prosodies in German reduced speech. Proc. 14th ICPhS, vol. 1, pp. 89–92, San Francisco 1999.
Kohler, K.J.: Investigating unscripted speech: implications for phonetics and phonology; in Festschrift für Björn Lindblom. Phonetica *57:* 85–95 (2000).
Kohler, K.J.: Articulatory dynamics of vowels and consonants in speech communication. J. Int. Phonet. Ass. *31:* 1–16 (2001a).
Kohler, K.J.: Plosive-related glottalization phenomena in read and spontaneous speech: a stød in German? In Grønnum, Rischel, To honour Eli Fischer-Jørgensen. Travaux du Cercle Linguistique de Copenhague, vol. 31, pp. 174–211 (Reitzel, Copenhagen 2001b).
Kohler, K.J.: The investigation of connected speech processes: theory, method, hypotheses and empirical data. Arbeitsber. Inst. Phonetik Univ. Kiel (AIPUK) *35:* 1–32 (2001c). http://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35/kkb.pdf
Kohler, K.J.: Phrase-level sound structures in French; in Barry, Pützer, Festschrift für Max Mangold zum 80. Geburtstag. PHONUS No. 6, pp. 129–157 (Institut für Phonetik, Saarbrücken 2002).
Kohler, K.J.: On the notion of 'fine phonetic detail': the case of speech reduction. Speech Reduction Workshop at MPI for Psycholinguistics, Nijmegen 2008. http://www.ipds.uni-kiel.de/kjk/pub_exx/kk2008_5/nijmegen_reduction.zip
Kohler, K.; Pätzold, M.; Simpson, A.: From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech (IPDS, Kiel 1995).
Lindblom, B.: Economy of speech gestures; in MacNeilage, The production of speech, pp. 217–245 (Springer, New York 1983).
Lindblom, B.: Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle, Marchal, Speech production and speech modelling, pp. 403–439 (Kluwer Academic Publishers, Dordrecht 1990).
Local, J.: Variable domains and variable relevance: interpreting phonetic exponents. J. Phonet. *31:* 321–339 (2003).
Macneilage, P.F.: The origin of speech (Oxford University Press, Oxford 2008).
Menzerath, P.; de Lacerda, A.: Koartikulation, Steuerung und Lautabgrenzung (Dümmlers, Berlin 1933).
Niebuhr, O.; Kohler, K.J.: Perception of phonetic detail in the identification of highly reduced words. J. Phonet. (in press).
Öhman, S.E.G.: Coarticulation in VCV utterances: spectrographic measurements. J. acoust. Soc. Am. *39:* 151–168 (1966).
Passy, P.: Étude sur les changements phonétiques et leurs caractères généraux (Librairie Firmin-Didot, Paris 1890).
Passy, P.: French phonetic reader (University of London Press, London 1929).
Pike, K.L.: Phonetics: a critical analysis of phonetic theory and a technique for the practical description of sounds (University of Michigan Press, Ann Arbor 1943).
Sachs, L.: Statistische Auswertungsmethoden (Springer, Berlin1972).
Simpson, A.P.: Casual speech rules and what the phonology of connected speech might really be like. Linguistics *30:* 535–548 (1992).
Welford, A.T.: Choice reaction time: basic concepts; in Welford, Reaction times, pp. 73–128 (Academic Press, New York 1980).
Wesener, T.: The phonetics of function words in German spontaneous speech. Arbeitsber. Inst. Phonetik Univ. Kiel (AIPUK) *34:* 327–377 (1999).
Wesener, T.: Some non-sequential phenomena in German function words. J. Int. Phonet. Ass. *31:* 17-27 (2001).
West, P.: Perception of distributed coarticulatory properties of English /l/ and /ɹ/. J. Phonet. *27:* 405–426 (2000).