ARTICULATORY PROSODIES IN GERMAN REDUCED SPEECH

Klaus J. Kohler
Institute of Phonetics and Digital Speech Processing, University of Kiel, Germany

ABSTRACT

Starting from spontaneous speech data of the Kiel Corpus, reduct-ion patterns of function words are described, which also incorporate more global articulatory prosodies, such as nasality, labiality and glottalization. The results of 4 perceptual experiments support the hypothesis that these long components of speech production are mapped onto perception. The discussion is also a contribution to a new paradigm for the analysis of non-lab and non-scripted speech.

1. PRODUCTION PATTERNS OF FUNCTION WORDS

1.1. Articulatory Fusion and Lexicalization

The phonetic form of function words in German shows great variability along an articulatory scale from strong elaboration to a high degree of reduction [7, 8], depending, among other factors, on prosody, especially sentence accent as well as prosodic grouping, and context of situation. In sequences of unstressed function words, reduction may result in their articulatory fusion, e.g., of prepositions with articles (a) or of auxiliary verbs with enclitic pronouns (b).

This may lead to the emergence of new lexical items, for example, in category (a), to monosyllabic "zum" [tsʊm] (*"to the"*) by the side of bisyllabic "zu dem", which varies along the reduction scale from [tsü dëm] to [tsʊ bm]. Although related historically and with regard to articulatory reduction, the mono- and bisyllabic phonetic forms today pattern differently in phrasal collocation due to the lexicalization of "zum", as in "er kam zum Schluß" (*"he came at/to the end"*) vs. "er kam zu dem Schluß, daß..." (*"he reached the conclusion that..."*). In the parallel case "mit dem" (*"with the"*), on the other hand, articulatory fusion to monosyllabic [mɪm], beyond the bisyllabic reduction scale from [mɪt dëm] to [mɪ bm], does not result in a new lexicalization.

Similarly, "haben/können/sind/sollen/wollen wir" *("have/can/are/shall/will we")*, in category (b), vary along the scales from [häb̈n vi]/[k̈n̈n vi]/[zɪnt vi]/[z̈ln vi]/[v̈ln vi] to [haṁ]/[kṁ]/[zɪṁ]/[z̈ṁ]/[v̈ṁ], resulting in a new inflect-ional paradigm of fused lexical items that are all bisyllabic. This reduction tendency is particularly strong in spontaneous speech.

1.2. Articulatory Prosodies of Nasality and Labiality

The Kiel Corpus [2] contains the example "nun wollen wir mal kucken" (*"now let's see"*) in the phonetic form [nṹ ṅ̽ ŋa¨k^{h}ʊkŋ] for unreduced [nün v̈ln vi mäl˙k^{h}ʊkŋ] [7, 9, 10]. It has strong nasalization across its first three syllables relating to syllable-final nasal consonants, which are reduced (deleted or shortened) in this hypo as against the hyper pronunciation. There is additional labiodentalization around the third syllable representing canonical [v] of wir. Other possible realizations are [nṹ (ŋ/m̃) ma¨k^{h}ʊkŋ] [10], where the apical gesture of the medial nasal is also eliminated or the consonant deleted altogether. So in these fusions of function words articulatory residues may persist as non-linear, suprasegmental features of syllables, reflecting, e.g., nasality or labiality that is no longer tied to specific segmental units.

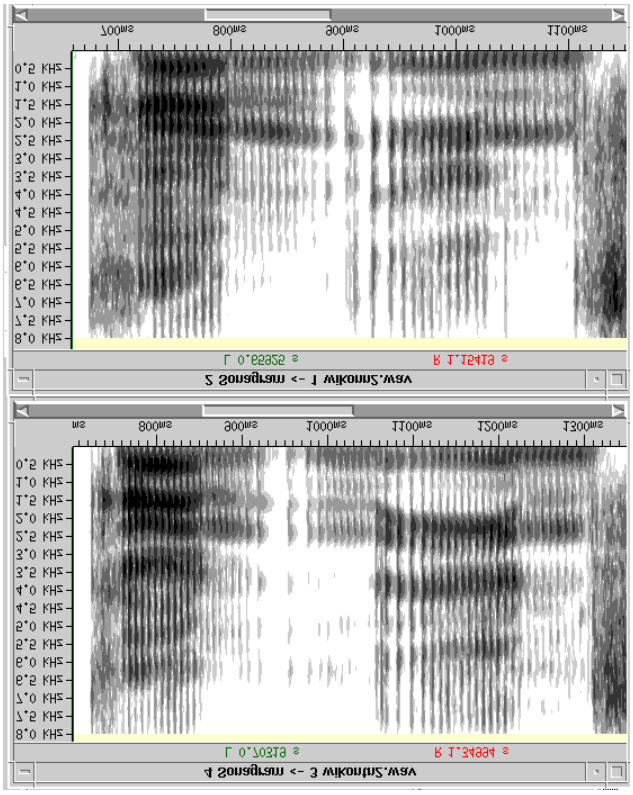
In those cases where the vowel in the first syllable bears a close acoustic relationship to the vowel in the second syllable, i.e. [i˙] [i˙] in [k̈ṁ]/[z̈ṁ]/[v̈ṁ], the reduction can go further to a nasalised monosyllabic realization [k̃()]/[z̃()]/[ṽ()], in, e.g., "k_nnen/sollen/wollen wir das machen" (*"can/shall/will we do this"*). Thus "nun wollen wir mal kucken" may also be expected to be realised as [nṹ() ma¨k^{h}ʊkŋ].

So nasalization may be the only articulatory parameter left to differentiate the production of "sollen wir das machen" [z̃ das ˈmaẍn] from "soll er das machen" [z̈ das ˈmaẍn] (*"is he to do it"*), which lacks it. The same nasal/oral dichotomy may apply to "sollen sie" [z̃zi] (*"are they to"*) vs. [z̈zi] (*"is she to"*).

1.3. Phonatory Prosody of Glottalization

Besides the articulatory prosodies of nasality and labiality, the verb-al paradigm also makes use of glottalization to differentiate "könn-ten/sollten/wollten wir/sie" (*"could/should/would we/they"*) vs. "können/sollen/wollen wir/sie". Instead of stopping the air stream for [t] by velic action in a nasal context, the velum may be lowered throughout, and the signalling of a break, as for a plosive, may then be achieved by glottal activity, in the extreme case by a glottal stop, but irregular glottal vibration is equally possible at any point during the nasal segment [5]. Thus [kn̰]/[z̈n̰]/[v̈n̰], [kn̰ zi]/[z̈n̰ zi]/[v̈n̰ zi] (alternatively with an additional syllabic modal-voice nasal after the glottalized part) are possible realizations.

Furthermore, the nasal stop articulation may again be replaced by syllable nasality overlaying at least the first vowel and the glottalization then associated with its final section, as in [k̰ zi]/[z̰ zi]/[v̰ zi]. That is less likely to occur in the context of "wir" because in more hyper production this type of glottalization is con-nected with an oral occlusion, and therefore presupposes a conson-antal gesture, as in the position before [zi]. But in [m̰], after a vowel, the elimination of a labial closing movement leads to vowel-internal glottalization. However, the consonantal link remains and the probability of occurrence increases if, within a frame of global syllable nasality, glottalization is coupled with a lip gesture into and out of an approximant stricture, e.g. in [z̃ṽ] or [z̃β̃].



The distinction between a consonantal and a vocalic base of glottalization plays an important role in German phonology. An example of the former is the replacement of plosives in a sonorant, especially nasal environment, as outlined above; the latter functions as a word-initial boundary marker. The different vocal tract reson-ances for the irregular glottal pulses in the two cases are illustrated in the spectrographic analysis of "wir könnten ihn fragen" (*"we could ask him"*) [vi knn̰n ḭn̰ frägŋ] in Figure 1. Glottalization to mark vowel onset may start in a preceding sonorant configur-ation, but this "overspill" is much shorter than the actual vocalic-base glottalization, e.g. in "wir können ihn fragen" (*"we can ask him"*) [vi knn ḭn̰ frägŋ] of Figure 2.

1.4. Reduction Rules

Reduction of function words in German exhibits patterns which can be formulated in the following rules by reference to accented strong citation form pronunciations:

(1) The degree of reduction depends on word class, morphol-ogical, syntactic and prosodic structures as well as speaking style. It is particularly high for articles and their combinations with prepositions as well as for enclitic sequences of auxiliary verbs and pronouns, in certain cases resulting in new lexicalizations.

(2) Diphthongs tend towards monophthongization, long vowels towards shortening and all vowels towards more central and mid positions. In extreme cases the result is [ʔ], or [ʔ] when phonological /r/ is involved: "ein" [ʔn], "der" [ʔd], "wir" [ʔv̈], "mit dem" [mɪd/tm], "zum" [tsm], "zur" [ts̈].

(3) The glottal word boundary marker of an initial vowel may be eliminated inside unaccented article + preposition and auxiliary + enclitic pronoun constructions: "auf einen/einem", "soll er".

(4) [ʔ], including the result of (2), may be deleted, e.g. "haben/können/sollen/wollen" [häbm]/[k̰nn/[z̈ln]/[v̈ln], "ein" [n], "mit dem" [mɪd/tm], "zum" [tsm]. Subsequent rules also apply to the segment sequences resulting from (4).

(5) Interconsonantal /t/ may be deleted: "sind wir" [zɪn v̈].

(6) Apical stop consonants are assimilated in place to follow-ing labials/dorsals, irrespective of word boundaries; apical nasals are also adjusted to preceding labials/dorsals within the same word: "haben" [häbm], "mit dem" [mɪb/pm], "können/sind/sollen/wollen wir" [k̰ŋ/mv̈]/[zɪŋ/mv̈]/[z̈lŋ/mv̈]/[v̈lŋ/mv̈].

(7) Final /l/ may be deleted, even before initial vowels of en-clitic words: "mal", "soll(en)", "solch", "welch", "will", "wollen".

(8) Velic closure in lenis plosives before nasals may be cut out and [ŋ/mv] integrated in a single bilabial nasal gesture [m]: "haben" [häm], "mit dem" [mɪm];[hamv̈] etc..

(9) The velic closing movement for a plosive in a nasal environment is relaxed and a prosody of irregular glottal activity produced instead: "könnten" [knn̰n], "sollten wir" [z̈m̰(mv̈)].

(10) A postvocalic closing movement for a nasal consonant is reduced or totally eliminated, with nasality spreading, particularly across the preceding vowel; mid to open diphthongs may be mono-phthongized: "sollen wir" [z̃ʊ̃][z̃][z̃], "sollten wir" [z̃ʊ̃].

2. PERCEPTION PATTERNS OF FUNCTION WORDS

2.1. A Hypothesis and a New Experimental Frame

Since the production patterns found in function words are an essent-ial feature of connected, especially spontaneous speech it must be assumed that they also play a fundamental role in speech percept-ion. The question thus is as to how listeners make use of phonetic parameters contained in reduced speech to restore the intended words and utterances, and what relevance should be attributed to the global prosodic features of nasalization, labialization, and glottalization, as well as to articulatory residues, over and above segmental information, for correct decoding of connected speech.

To test this hypothesis we need a new type of data in our perception experiments, compared with the traditional paradigm, which uses very simple stimuli of syllable or word size, often of a nonsense word type, within a standard metalinguistic sentence frame, systematically varying acoustic parameters in speech syn-thesis. The Haskins experiments on VOT and second formant transitions are classic examples. The aim of such perception tests is to gain insight into the perceptual relevance of specific parameter values for phoneme perception in word citation forms.

None of these heavy constraints apply to the phenomena at issue at the utterance level. We first of all need speech data of at least sentence size in a natural and meaningful context because it is only there that the reductions occur and can be tested perceptually. Secondly, the acoustic quality must be completely natural, so only high-level speech synthesis or very careful time-domain splicing are feasible tools for signal manipulation. Thirdly, the auditory test stimuli must be modelled on production data found in large con-nected speech data bases, and they should be convincingly reprod-uced and systematically varied in natural production by competent, phonetically trained native speakers for subsequent processing and parameter variation according to new experimental designs. Finally, the perceptual test will have to put articulatory prosodies rather than segment-type phonemes in focus [6].

Such a test frame was implemented in the following steps:

- The spontaneous speech example discussed in 1.2 as well as glottalization data from the Kiel Corpus were the start.
- The phrases (a) "soll er", "soll(t)en wir"; (b) "soll sie", "soll(t)en sie"; (c) "wir könn(t)en ihn"; (d) "die könn(t)en (wir) uns" were put in the utterance frames (a,b)"Was meint ihr?__das machen?" (*"What do you think? __do it?"*); (c) "__fragen." (*"__ask him."*); (d) "__abholen." (*"__collect them (for ourselves)."*). selected for processing.
- The data were then analysed in *xassp* [3] and systematic splicing was applied to create test stimuli for 4 listening tests: (a) *soller* (b) *sollsie*, (c) *konnfra*, (d) *konnab*.
- The author produced each of these utterances with a broad array of phrase realizations from hyper to hypo along the scales found in connected speech data. The recordings were auditorily screened and the most convincing rendering of each utterance, phrase and reduction type

- Each test stimulus, preceded by a 50ms sine warning tone and followed by a 4s pause, was copied 10 times. The stimuli were randomised and copied onto 4 separate analog test tapes.
- Questionnaires were prepared with three alternatives in (a) and (b) ("soll er", "sollen wir", "sollten wir"; "soll sie", "sollen sie", "sollten sie") and two alternatives in (c) and (d) ("können", "könnten") for forced choice answers.
- All 4 tests were administered in a sound-treated room via loudspeaker, first to a group of phonetics students in two sess-ions in the order (a), (b) and (c), (d) and then to another group of phonetically naive students in one session. The complete test with introduction and breaks lasted about 75 mins.

2.2. Prosodies of Nasalization and Glottalization: Exp. *soller*

2.2.1. Stimulus description and results. All stimuli for Exps. *soller* and *sollsie* have identical frames "s__ das machen", taken from the selected production "soll sie das machen" [z̈ˡ zi das̈ maẍn]. For Exp. *soller* the following excerpts were spliced into this frame: "(s)oll er" [l̈˙] (***ser1***), "(s)oll er" [l̈˙] (***ser2***), "(s)ollen wir" [lŋ̈˙] (***ser3***), "(s)ollen wir" [l̈˙] (***ser4***), "(s)ollten wir" [lʔ mj̈] (***ser7*)**, all from the selected stimulus set "soll er/soll(t)en wir das machen"**.**

Then signal manipulations were carried out. (a) One period was removed from the centre of [l̈] in *ser4* to give this stimulus the same duration as *ser2*. (b) [l̈] from natural *ssi4* was spliced into the frame (***ser5***). The *ssi4* vowel was lengthened by duplicating central periods to give it the same duration as [l̈˙] in *ser4* (***ser6***). (c) Glot-talization pulses, excerpted from natural "(s)ollten wir" [ʔ mj̈], were spliced into the vowel centre of *ser4* and duplicated (***ser8***). (d) In *ser8* the second half of the glottalized section as well as 4 periods of the following modal voice were replaced by the signal segment [ŋ] of the source stimulus in (c) (***ser9***).

ser1,3,7 represent less reduced and therefore clear cases of "soll er", "sollen wir", "sollten wir", respectively. *ser2,4* are con-trasted by the absence/presence of a nasal prosody to differentiate "soll er" from "sollen wir". *ser5,6* have a short vs. long nasalized monophthong instead of the diphthong. *ser8,9* introduce a prosody of glottalization into a global nasal prosody without and with labiodentalization. The results of the listening test are in Table 1.

2.2.2. Discussion.

- The anchors *ser1,3,7* are uniquely identified.
- In *ser2,4* a nasal prosody across the diphthong can signal the "soll er"/"sollen wir" distinction, but no longer uniquely.
- As to the nasal monophthongs in *ser5,6*, the two groups behave differently: in Gr1 "soll er" dominates, in Gr2 "sollen wir", with fewer "sollen wir" for the short vowel than for the long one in both groups. The two groups rated the perceived nasality differently, presumably either as a speaker's voice quality or as a linguistic differentiator. One of the Gr1 subjects commented after the test that he had been uncertain as to these interpretations. This coincides with the fact that Gr1 were the phonetics students who knew my voice, which is characterised by slight nasality, whereas Gr2 were outsiders. But the quality and duration of the nasalized vowel were further factors.
- The introduction of glottalization into the nasalized vowel in *ser8,9* changes judgement to predominant "sollten wir" in both groups, but more so in the case of added labiodentalization.

	soll er		sollen wir		sollten wir	
Stim	Gr1	Gr2	Gr1	Gr2	Gr1	Gr2
ser1	120	100	0	0	0	0
ser2	99	93	21	7	0	0
ser3	0	0	120	100	0	0
ser4	27	14	89	83	4	3
ser5	95	37	25	62	0	1
ser6	82	28	37	71	1	1
ser7	0	0	0	0	120	100
ser8	11	6	6	22	103	72
ser9	2	0	3	5	115	95

Table 1. Absolute frequencies in the 3 categories of Exp. *soller*
Group 1 N=120, Group 2 N=100

- In *ssi2,4* the nasal prosody can no longer signal the "soll sie"/"sollen sie" contrast. *ssi2* is uniquely, *ssi4* predominantly identified as "soll sie" by both groups. Here the linguistic interpretation of nasality was probably difficult due to the short duration of the vowel (150ms vs. 220ms) and to insuffic-ient information from the situational context.
- Glottalization in the nasalized vowel of *ssi6,7* changes judge-ment to predominant "sollten sie" in Gr1, as expected, but in Gr2 there is large number of "soll sie" and for *ssi7* also of "sollen sie" responses, without a ready explanation.

2.3. Prosodies of Nasalization and Glottalization: Exp. sollsie

2.3.1. Stimulus description and results. For Exp. *sollsie* the following excerpts were spliced into the frame of 2.2.1: "(s)oll sie" [l̈ zi] (***ssi1***), "(s)ollen sie" [l̈n zi] (***ssi3***), "(s)ollen sie" [l̈ zi] (***ssi4***). "(s)ollten sie" [lʔn zi] (***ssi5*)**, all from the selected stimulus set "soll sie/soll(t)en sie das machen"**.**

Then signal manipulations were carried out. (a)The weak [ˡ] in the base stimulus of 2.2.1 was removed and compensated for by duplicating every second period within the vowel (***ssi2***). (b) Glottal-ization pulses, excerpted from natural "(s)ollten wir" [ʔ mj̈], were spliced in at the end of the vowel of *ssi4* and duplicated (***ssi6***). (c) Another stimulus was derived from *ssi6* by removing 3 periods of modal voice to get glottalization closer to the consonant [z] (***ssi7***).

ssi1,3,5 represent less reduced and therefore clear cases of "soll sie", "sollen sie", "sollten sie", respectively. *ssi2,4* are con-trasted by the absence/presence of a nasal prosody to differentiate "soll sie" from "sollen sie". *ssi6,7* introduce a prosody of glottaliz-ation into a global nasal prosody at different distances from the consonantal gesture. The results of the listening test are in Table 2.

	soll sie		sollen sie		sollten sie	
Stim	Gr1	Gr2	Gr1	Gr2	Gr1	Gr2
ssi1	120	100	0	0	0	0
ssi2	120	100	0	0	0	0
ssi3	3	1	117	99	0	0
ssi4	91	63	29	37	0	0
ssi5	0	0	2	0	118	100
ssi6	0	24	3	7	117	69
ssi7	2	32	3	17	115	51

Table 2. Absolute frequencies in the 3 categories of Exp. *sollsie*
Group 1 N=120, Group 2 N=100

2.3.2. Discussion.

- The anchors *ssi1,3,5* are uniquely identified.

2.4. Timing and Resonance of Glottalization: Exps. *konnfr/ab*

2.4.1. Stimulus description and results. For Exp. *konnfr* the following stimulus selection and processing was done: (a) original "können ihn" with /i/ glottalization (Figure 2) (***kfr1***), and original "könnten" with /n/ glottalization, put in the sentence frame of *kfr1* (***kfr9***); (b) deletion of glottalization in *kfr1* (***kfr2***), combined with compensatory lengthening of modal /n/ (***kfr3***); (c) replacement of the glottalized section in *kfr1* by /n/ glottalization from original "könnten" of Figure 1 (***kfr4***), complementary glottal lengthening and voice shortening - partial/total (***kfr5,6***); (d) lengthening of /i/ glottalization in *kfr1* (***kfr7***), preceded by equal-length /n/ glottal-ization from the stimulus in Figure 1 (***kfr8***). Stimuli *kfr1,3,4,5,6,7*

have approximately the same duration to the onset of modal /i/.

In the construction of Exp. *konnab* original "können" was selected (***kab1***) and used as the base for (a) /n/ lengthening (***kab2***) and shortening (***kab3***), (b) splicing /n/ glottalization from original "könnten" into the second or first half or the centre or the total duration of /n/ (***kab4,5,6,7***). Original "könnten", spliced into the same sentence frame as the other items, was included as a reference stimulus (***kab8***). The results of the listening tests are in Table 3.

Stim	können	könnten	Stim	können	könnten
kfr1	222	8	kab1	230	0
kfr2	230	0	kab2	230	0
kfr3	230	0	kab3	229	1
kfr4	3	227	kab4	3	227
kfr5	0	230	kab5	1	229
kfr6	2	228	kab6	0	230
kfr7	65	165	kab7	1	229
kfr8	220	10	kab8	0	230
kfr9	0	230			

Table 3. Absolute frequencies in the 2 categories of Exps. *konnfra* (left) and *konnab* (right); combined groups N=230

2.4.2. Discussion.

- There were no differences between the two groups.
- All stimuli that contain /n/ glottalization only, irrespective of its position in the nasal and its extension, are uniquely identified as "könnten": *kfr4,5,6, kab4,5,6,7,8*.
- All stimuli that have only modal voice or /i/ glottalization are uniquely identified as "können": *kfr2,3,8, kab1,2,3*. This also applies to *kfr1*, where there is a "spillover" of glottalization into /n/, which is however small in relation to the one in /i/.
- When contiguous /n/ and /i/ glottalization are both lengthened judgement is no longer unique, but "könnten" predominates.

3. CONCLUSION

The analysis of German function word production in connected and spontaneous speech has shown the importance of global parameters in articulatory reduction. Articulatory components become dissoc-iated from segmental entities of speech sound size, manifesting themselves at highly variable points in time and with highly flexible extension, usually in larger units of at least syllable size. Thus glot-talization instead of velic elevation for plosives in a nasal environ-ment may occur anywhere during the nasal context and with a large range of durations up to the total length of the latter.The only constraints are the incidence of the phenomenon and its link with consonantal vocal tract resonance. If there is a "spillover" into a vocalic configuration it is relatively short.

Nasalization may become a feature of a syllable or a whole syllable chain and not be tied to a delimitable nasal consonant. The same applies to labi(odent)alization. Both nasal and labi(odent)al consonants may be absent as separate units as long as the nasal and labial gestures are integrated in the total articulatory complex. This results in articulatory residues in the fusion of words.

These patterns of speech production find their parallel in speech perception. The temporal indeterminacy in the production of glottalization in nasal consonants is mapped onto a perceptual insensitiveness to this temporal variability. Similarly, syllable nasal-ization as a residue of nasal consonant deletion becomes a cue for reduced function word perception.

The long articulatory components and their perceptual counter-parts may be regarded as prosodies in the Firthisan sense [1]. To deal with them adequately it is mandatory to transcend the tradition-al paradigm of phonology in two ways, by leaving word phonology and moving on to phrase level phonology [6], and by abandoning the srtrictly linear frame-work in favour of 'complementary phonology' [4], which attributes a theoretical status to articulatory prosodies by the side of phonemic segments in an 'as well as' approach.

Investigations of phrase level phenomena in general and of speech reduction in particular also require different kinds of data bases from those collected with the predominant concern for quest-ions of individual word pronunciations. As regards speech product-ion, large corpora of dialogue interactions, of a more natural, spont-aneous nature than read and lab speech, have been collected and analysed. This development was largely triggered by speech tech-nology projects. These data analyses have made it possible to provide detailed descriptions of reduction patterns in German.

The same reorientation towards more natural, phrase level data is also necessary in perception research. It demands a new meth-odology that introduces not only natural speech data but also a strong focus on context of situation into the experimental design.

REFERENCES

[1] Firth, J. R. 1948. Sounds and prosodies. *Transactions of the Philological Society* , 127-152.

[2] IPDS. 1994-1997. *The Kiel Corpus of Spontaneous Speech.* Vols. 1-3. CD-ROM#2-4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.

[3] IPDS**.** 1997. *xassp* User's Manual (*A*dvanced *S*peech *S*ignal *P*rocessor under the *X* Window System)**.** *AIPUK* 32**,** 31-115.

[4] Kohler, K. J. 1994. Complementary phonology. A theoretical frame for phrase level phonology. *Proc. CSLP94*. Yokohama.

[5] Kohler, K. J. 1996. Glottal stop and glottalization. A prosody in European languages. *AIPUK* 30, 207-216.

[6] Kohler, K. J. 1996. Developing a research paradigm for sound patterns of connected speech in the languages of the world. *AIPUK* 31, 227-233.

[7] Kohler, K. J. 1998. The phonetic manifestation of words in spontaneous speech. In Duez, D. (ed.), *Proc. from the Esca Workshop on Sound Patterns of Spontaneous Speech.* La Baume-les-Aix.

[8] Rehor, C. 1996. Phonetische Realisierung von Funktionswörtern im Deutschen. *AIPUK*, 30, 1-113.

[9] Simpson, A. P. and M. Pätzold (eds.). 1996. *Sound Patterns of Connected Speech: Descriptions, Models and Explanations. AIPUK* 31.

[10] URL: www.ipds.uni-kiel.de/examples.html