Articulatory reduction in fluent speech
A pilot study on syllabic [n] in spoken Standard German

Bernd Pompino-Marschall

Zentrum für Allgemeine Sprachwissenschaft

Jägerstr. 10/11
D-10117 Berlin
Germany

SUMMARY

The paper reports of an acoustic and perceptual study on reduction of the German word final syllable *'_en'* [@n][1] in contrast to single word final [n]. Three utterances of *'ein'* and *'einen'* were cut from the same sentence utterances of four different speakers and analysed according to their segment durations. In a perception experiment these utterances had to be identified. Statistical analyses revealed that the perception of *'einen'* is mainly dependent on the duration of the final [n] segment itself. Detailed analysis of the results however showed that there may be quite complex articulatory reorganization resulting in the perception of the so called syllabic [n].

INTRODUCTION

Processes of reduction in fluent standard German have onlyoccasionally been studied (for more recent empirical studies see Wesenick 1994, Kohler 1996, for modelling Kohler 1990, 1992, Kröger 1993). The present paper describes an acoustic/auditory pilot study of schwa elision in nasal context resulting in the so called syllabic [n]. An identification test was run to determine the acoustic cues for the perception of syllabic [n].

STIMULI

As stimuli utterances of the words *'ein'* [QaIn] (German indefinite article, masculine, nominative case) and *'einen'* [QaIn@n] (indefinite article, accusative case masculine) from the PHONDAT 2 (cf. Pompino-Marschall 1992) Kiel corpus of read speech (IPDS 1995, Kohler 1994) were chosen. The utterances were selected to represent quite different forms of reductions of the *'einen'* items according to the original Kiel segmentation/transcription (IPDS 1995, Kohler 1994), i.e. items with clear as well as unclear segment boundaries between the single [n] instances and items with further [n] reduction. This selection resulted in the instances of the sentence utterances

Wann geht bitte ein Zug nach Hamburg mit möglichst wenig, ja, Bahnhöfen? (e068)

Was ist die billigste Möglichkeit für ein Bahnticket von hier nach Kiel? (e083)

Wann fährt bitte ein Zug von Regensburg nach Dortmund am Montag, am Montagmorgen? (e097)

Gibt es einen Zug, der in Köln länger hält und nach Dortmund weiterfährt? (e076)

[1] Throughout the paper transcriptions are given in SAMPA (Wells et al. 1992) notation.

Gibt es einen Nachtzug, der am dreiundzwanzigsten Zwölften morgens in Ulm ankommt? (s055)

Gibt es einen Zug, der am dreiundzwanzigsten Zwölften um zirka zwanzig Uhr in Oldenburg ist? (s063)[2]

produced by four speakers (3 male: dlm, hpt, uga; 1 female: rtd[2]). All speakers had a standard German pronunciation (northern provenance).

The stimuli for the perception experiment were cut from the sentence utterances according to the original Kiel segmentation (i.e. from the beginning of the (glottalized) diphthong [aI] till the end of the last segmented/labelled [n] of the indefinite article. In figures 1 to 4 the audio signals, sonagrams, RMS amplitude and fundamental frequency contours of representative items of *'ein'* and *'einen'* are depicted for the four speakers.[3] Table I lists the durations of the single segments according to the Kiel segmentation (reanalysed with respect to the duration of the glottalized segment).

Table I: Segment durations (in ms) of the selected utterances

		item					
		'ein'			*'einen'*		
speaker		e068	e083	e097	e076	s055	s063
dlm	q	-	-	46	-	-	48
	al	68	40	37	95	77	64
	n	39	53	47	22	52	42
					n 17	-*	%22
hpt	q	67	-	62	86	68	39
	al	69	43	40	79	62	150
	n	31	61	38	50	47	27
					n -	%34*	26
rtd	q	13	42	63	88	87	57
	al	67	42	46	71	75	83
	n	29	69	43	26	61	81
					n %44	-*	-
uga	q	-	-	-	-	23	-
	al	88	70	94	103	83	112
	n	33	69	50	35	55	40
					n %42	55	41

% denotes a not clearly definable segment boundary,
items marked with an asterisk are followed by a not clearly segementable word initial [n].

[2] The canonical pronunciation of the sentences in SAMPA is given in the appendix.

[3] The analyses as well as the preparation of the test tape were conducted with the help of the Signalyze 3.11 software for Apple MacIntosh.

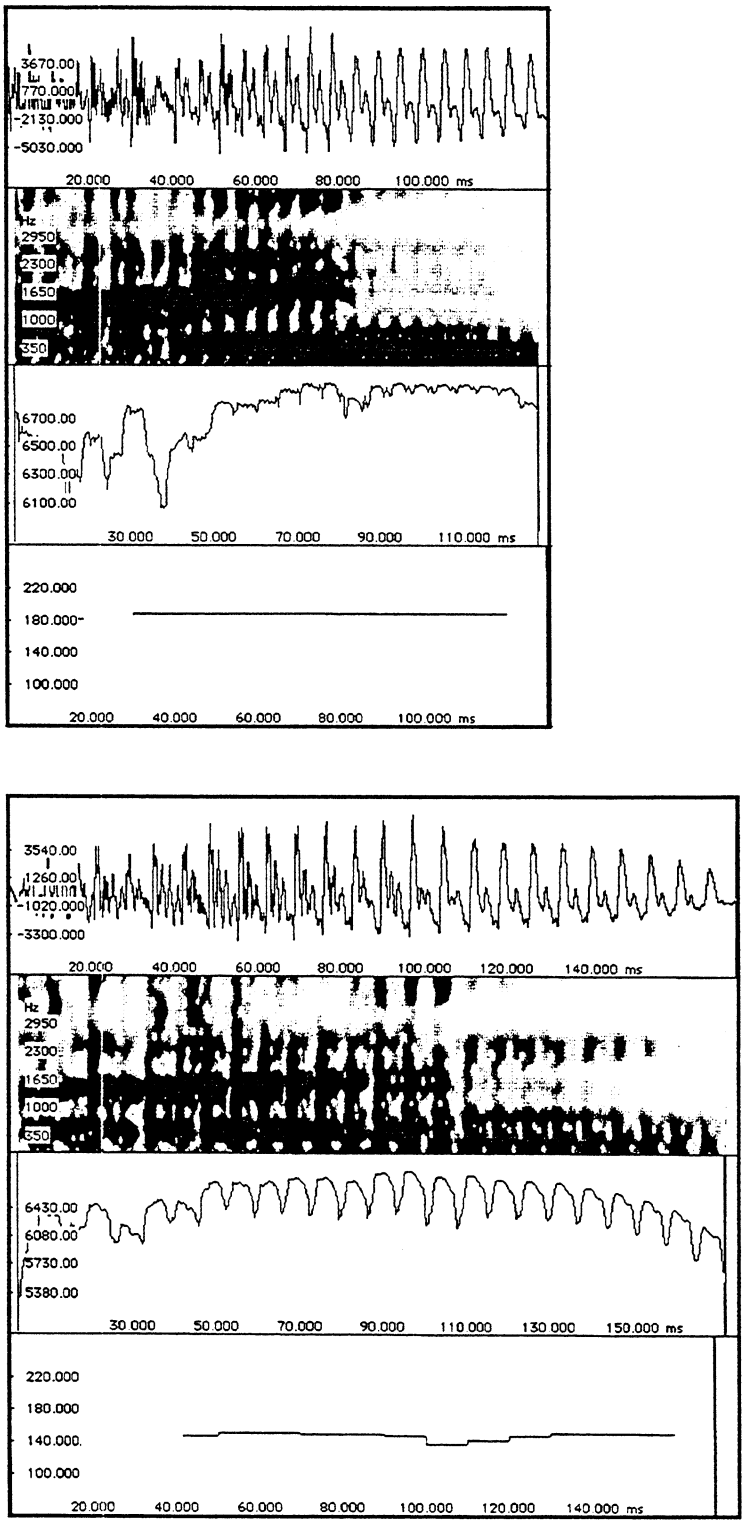
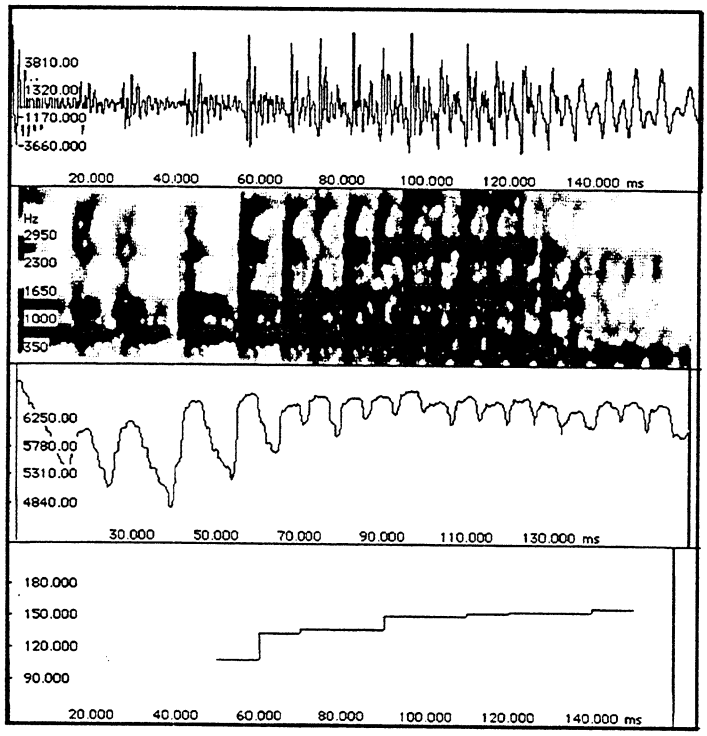


Fig. 1: Examples of utterances of *'ein'* (top; e097) and *'einen'* (bottom; s063) for speaker dlm; first row: audio signal; second row: wide band (200Hz) sonagram; third row: RMS amplitude (10 ms window); fourth row: fundamental frequency.



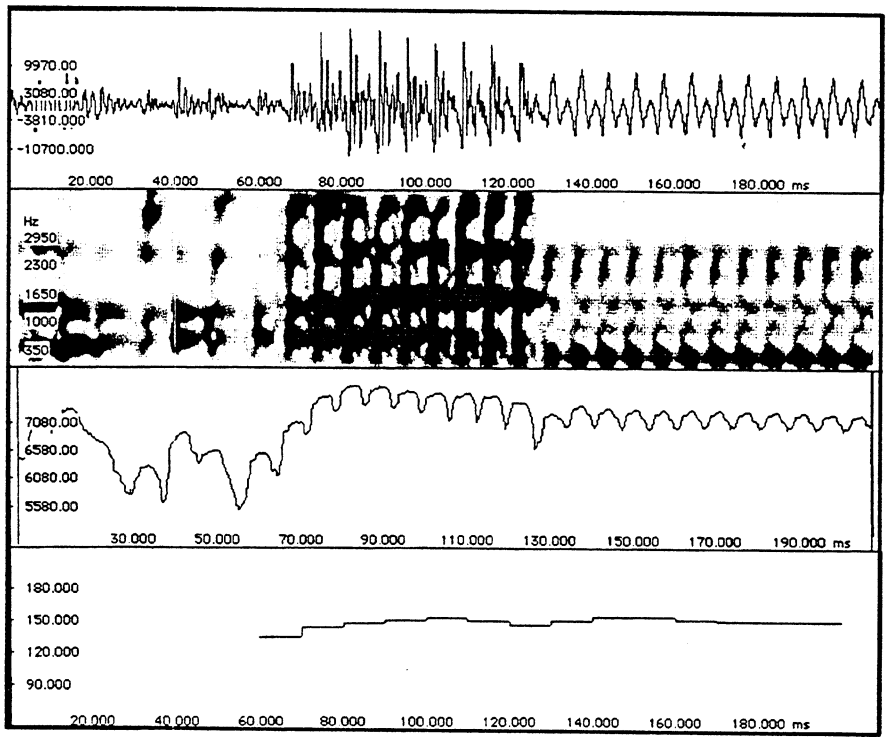


Fig. 2: Examples of utterances of *'ein'* (top; e068) and *'einen'* (bottom; s055) for speaker hpt (format as in figure 1).

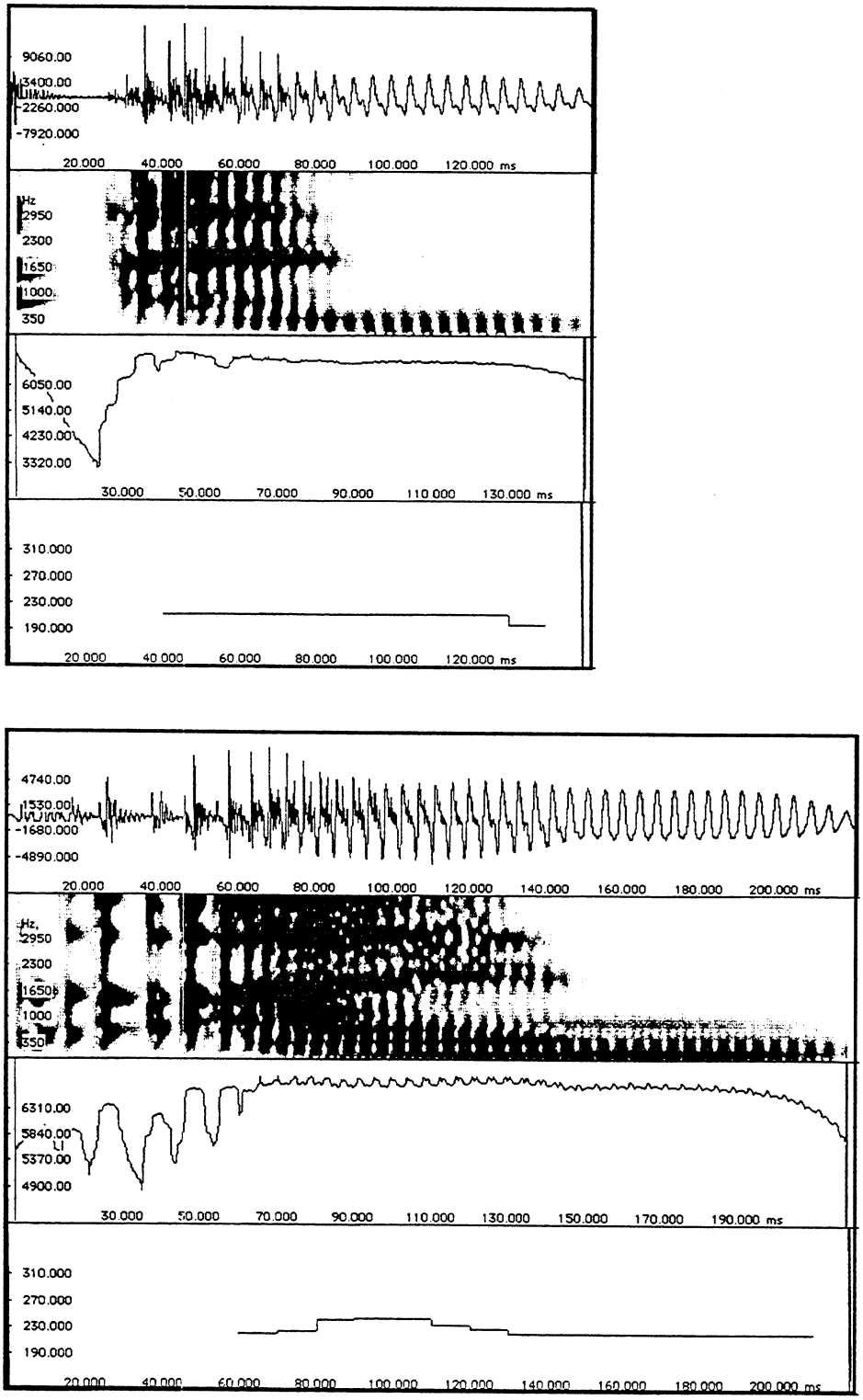
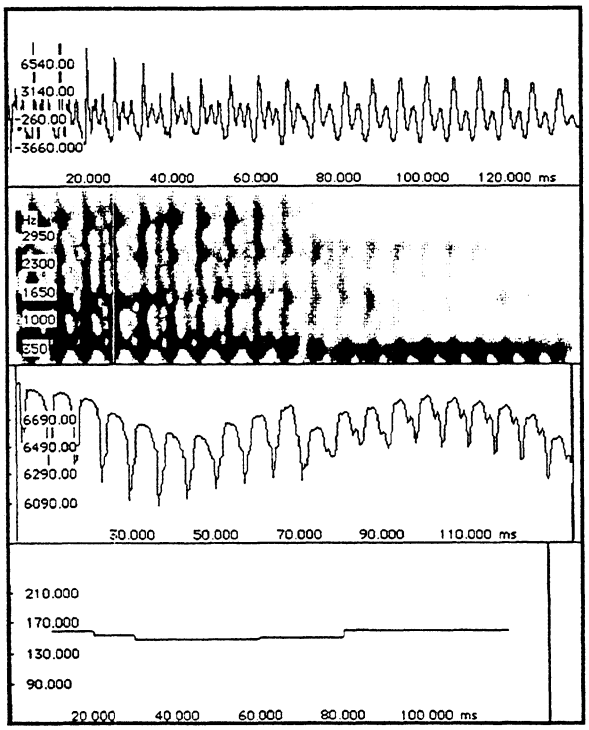


Fig. 3: Examples of utterances of *'ein'* (top; e083) and *'einen'* (bottom; s063) for speaker rtd (format as in figure 1).



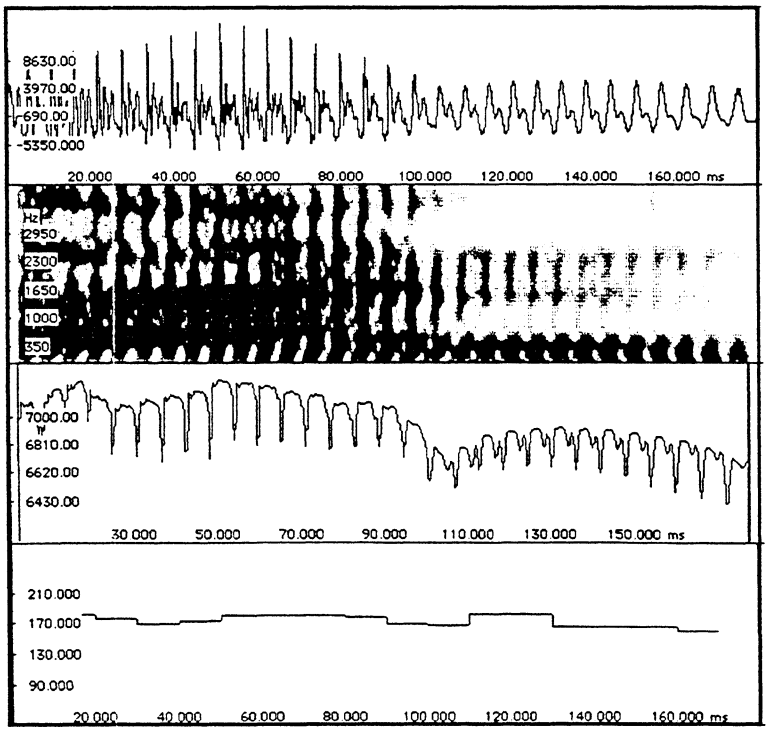


Fig.4: Examples of utterances of *'ein'* (top; e083) and *'einen'* (bottom; e076) for speaker uga (format as in figure 1).

EXPERIMENTAL PROCEDURE

The isolated *'ein'/'einen'* utterances were presented ten times in quasi randomized order to ten subjects (5 male, 5 female) who had to decide whether they heard *'ein'* or *'einen'*. The subjects listened to the stimuli via headphones at a comfortable loudness level at an interstimulus interval of about 3 s and a longer pause after blocks of 12 trials. They had to mark their choices on a prepared answer sheet.

RESULTS

The overall categorical responses to the single items is depicted in figure 5, the responses of the single subjects are listed in Table II.

Figure 5 clearly shows differences in the identification of the test items of different speakers and of single items as well. In Table II further differences in the responses of single listeners can be found (cf. below).

The *'ein'* utterances were identified correctly at 80.17 percent, the reduced *'einen'* utterances at only 66.08 percent. The latter is clearly due to the poor recognition of the speaker's dlm *'einen'* items; for the other speakers we get an identification score of 77% (again mainly resulting from difficulties with the single item e083 of speaker rtd; this item excluded: 81%) for *'ein'* and 80% for *'einen'*.

DISCUSSION

General observations

When cutting the test items from the utterances it was recognized that the glottalized portion of the diphthong [aI] often results in the perceptual illusion of a prefixed unstressed [d@] syllable. For the further analysis of segmental durations the utterances were therefore reanalysed according to the duration of glottalization as already shown in Table I.

Acoustic segment durations

Differences in segment durations between items of *'ein'* and *'einen'* are observable for the duration of the final [n] and the duration of the nonglottalized part of the diphthong [aI]. Overall, the difference in [n] duration - *'ein'*: 58.7 ms (19.89) vs. *'einen'*: 87.8 (24.60)[4] - is clearly significant ($t = 3.121$; $p < .01$): *'einen'* clearly has longer [n] durations. The different speakers show different lengthening of [n]: speaker dlm exhibits an only marginal lengthening of 12% whereas the speakers hpt, rtd, and uga show lengthenings of 42, 50, and 76% respectively. The diphthong [aI] - *'ein'*: 58.7 (19.89); *'einen'*: 87.8 (24.60) - is also significantly longer for *'einen'* ($t = 3.143$; $p < .01$). The different speakers again show different lengthening ratios not paralleling those for [n]. Clearly, the different context in sentence e083 - [fy:6 QaIn] - resulted in an extremely short diphthong in these items.[5] But even when these items were excluded from statistical analysis a significant difference in diphthong duration between *'ein'* and *'einen'* remained ($t = 2.362$; $p < .025$).

---

[4] Means and standard deviations (in brackets) in ms.

[5] In colloquial German there would even be the possibility of a total elision - [fy:6 n].

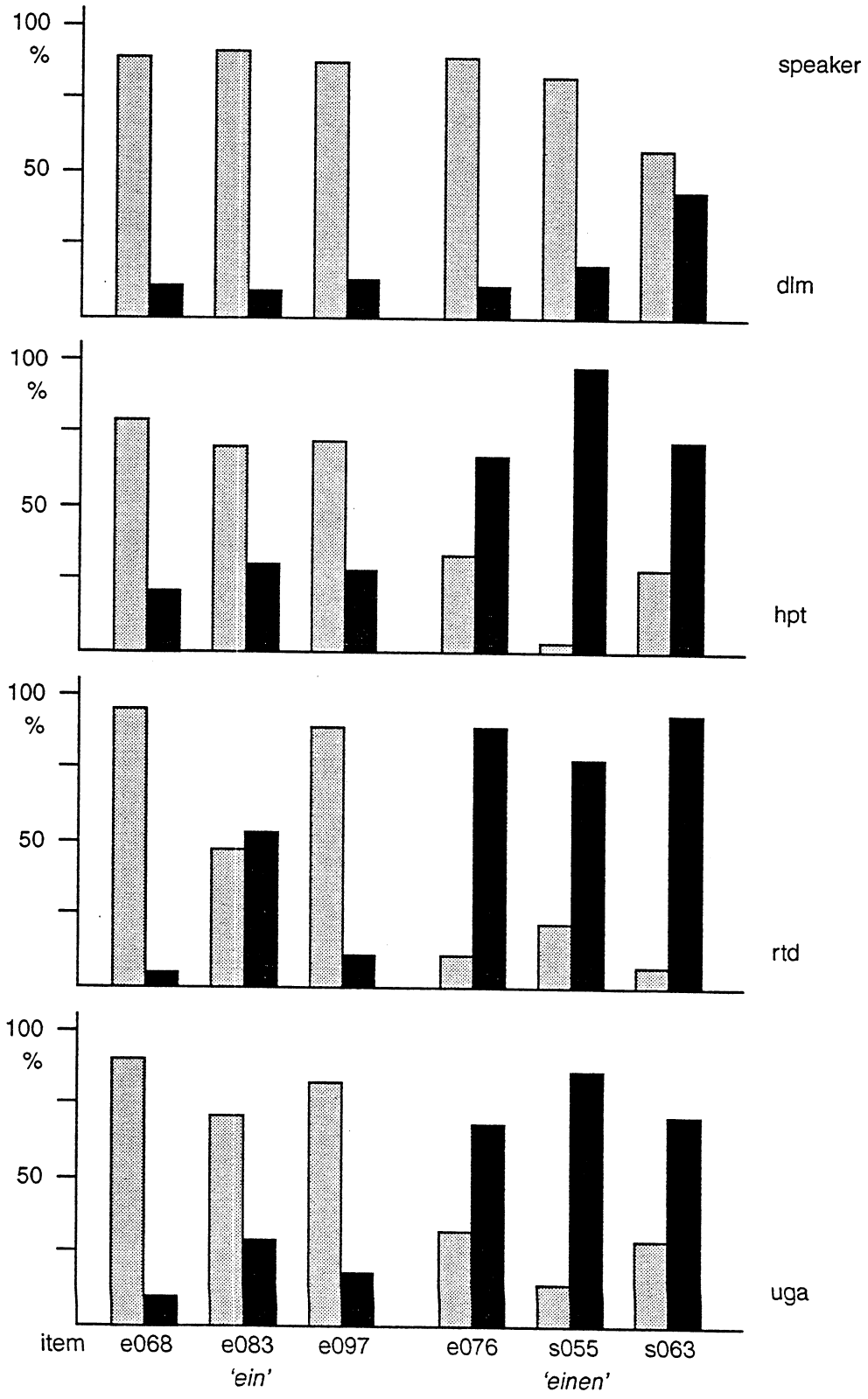


Fig. 5: Overall identification scores in percent; gray: *'ein'*, black: *'einen'*.

Table II: Response frequencies of *'ein'* (left columns) and *'einen'* (right columns) for the single subjects (single rows) and mean percentage (rows labelled sum).

speaker		10	-	9	1	10	-	10	-	10	-	6	4
		9	1	10	-	10	-	10	-	2	8	6	4
		10	-	10	-	10	-	10	-	10	-	7	3
		10	-	10	-	9	1	9	1	9	1	6	4
dlm		7	3	6	4	5	5	8	2	5	5	4	6
		10	-	10	-	8	2	8	2	10	-	2	8
		10	-	8	2	10	-	10	-	10	-	9	1
		8	2	9	1	7	3	8	2	9	1	2	8
		8	2	9	1	9	1	9	1	8	2	7	3
		7	3	10	-	9	1	7	3	9	1	8	2
	sum	89	11	91	9	87	13	89	11	82	18	57	43
		9	1	9	1	6	4	1	9	-	10	-	10
		10	-	10	-	10	-	9	1	-	10	10	-
		7	3	10	-	9	1	2	8	1	9	-	10
		5	5	8	2	5	5	-	10	-	10	1	9
hpt		4	6	1	9	3	7	3	7	-	10	1	9
		10	-	6	4	10	-	5	5	-	10	2	8
		8	2	5	5	8	2	10	-	1	9	9	1
		7	3	2	8	7	3	-	10	-	10	1	9
		10	-	9	1	6	4	3	7	-	10	3	7
		9	1	10	-	8	2	-	10	1	9	1	9
	sum	79	21	70	30	72	28	33	67	3	97	28	72
		10	-	9	1	8	2	-	10	-	10	-	10
		9	1	2	8	10	-	2	8	5	5	-	10
		10	-	7	3	9	1	-	10	4	6	-	10
		10	-	10	-	10	-	-	10	-	10	-	10
rtd		7	3	4	6	8	2	1	9	5	5	2	8
		10	-	3	7	10	-	1	9	-	10	1	9
		10	-	2	8	9	1	2	8	4	6	2	8
		10	-	-	10	10	-	1	9	2	8	-	10
		10	-	6	4	9	1	3	7	1	9	1	9
		9	1	4	6	6	4	1	9	1	9	1	9
	sum	95	5	47	53	89	11	11	89	22	78	7	93
		8	2	8	2	10	-	2	8	-	10	1	9
		10	-	2	8	7	3	2	8	2	8	3	7
		10	-	10	-	10	-	5	5	4	6	7	3
		10	-	9	1	9	1	6	4	-	10	3	7
uga		5	5	1	9	3	7	3	7	-	10	2	8
		10	-	7	3	8	2	-	10	-	10	1	9
		10	-	9	1	10	-	7	3	3	7	8	2
		8	2	8	2	7	3	-	10	-	10	1	9
		10	-	9	1	9	1	5	5	3	7	3	7
		9	1	8	2	9	1	2	8	2	8	-	10
	sum	90	10	71	29	82	18	32	68	14	86	29	71
	item	e068		e083		e097		e076		s055		s063	

Acoustic cues for syllabic [n]

Since no clear indications for a syllabic [n] - partly contrary to the expectations according to the original Kiel segmentation/transcription - could be found in the RMS amplitude contour as well as the fundamental frequency contour (cf. figures 1 to 4), the correlations between different segment durations and response frequencies were calculated.

Clearly, here the correlation between the duration of the (complex) [n] segment and the frequency of *'einen'* responses wins with $r = .7878$ ($p < .001$). Contrary to the primary expectation the correlation between the ratio of the duration of the nonglottalized vocalic portion and [n] failed significance ($r = -.4916$; n.s.).[6]

Specific observations

Some results concerning the different speakers, indiviual test items as well as single listeners require some further attention. The overall poor recognition (cf. figure 5 & table II) of the items of speaker dlm even when the original Kiel segmentation/transcription would predict otherwise (cf. table I, especially dlme076) may be interpreted as due to this speaker's weak variation in [n] duration (cf. above) between *'ein'* and *'einen'*. A clear effect of [n] duration also seems to be the cause for the poor identification of rtde083: here the [n] duration of *'ein'* is in the same range as in this speaker's *'einen'* items. For some listeners this [n] lengthening in e083 also results in categorical reversals for speaker uga.

The most interesting result is the categorical reversal in the identification of *'einen'* (s063) of speaker hpt by listeners 2 and 7. When inspecting this item (cf. figure 6) it can be seen that there is no obvious [n] lengthening but that the vocalic portion consists of two parts divided by a decrease in amplitude: an [a] part of about 49 ms and a diphthongal part. If the former is cut from the quite normal sounding original utterance as visualized in figure 6 a clearly ungrammatical utterance will result in that it is clearly perceived as *'Gibt es ein* (nom.) *Zug ...'* instead of *'Gibt es einen* (acc.) *Zug ...'*. What we get here is not a syllabic [n] but a bisyllabic indefinite article. Without the sentence context however, this bisyllabicity seemed to be no cue for these two listeners.

General discussion

The main cue for the perception of syllabic [n] therefore seems to be the duration of the [n] segment itself. If there is a further dependency on adjacent vowel duration which cannot finally be excluded from the present results this vowel duration is the duration of the nonglottalized vowel portion alone.

In the present data no indication of a really syllabic [n] could be found. In connected speech however, bisyllabicity on the other hand may not be cued by the [n] segment itself but also by an articulatory reorganization of adjacent segments. A prototypical example of such reorganization was recently reported by Tillmann (1995): The - by German hearers - clearly perceived utterance of the word *'und'* [QUnt] in the sentence *'Ich will nach Hamburg* [pause] *<u>und</u> möchte dort etwa gegen sieben Uhr abends ankommen'*[7] was only a glottalized pitch rise on the following [m] resulting in a syllabic [m] before the following word. A similar reorganization in articulatory timing resulted in the bisyllabicity of the vocalic nucleus of the *'einen'* item of hpts063 discussed above.

---

[6] No correlations at all resulted when the glottalized portion was included in this proportion.

[7] Also from the PHONDAT 2 corpus; Munich speaker.

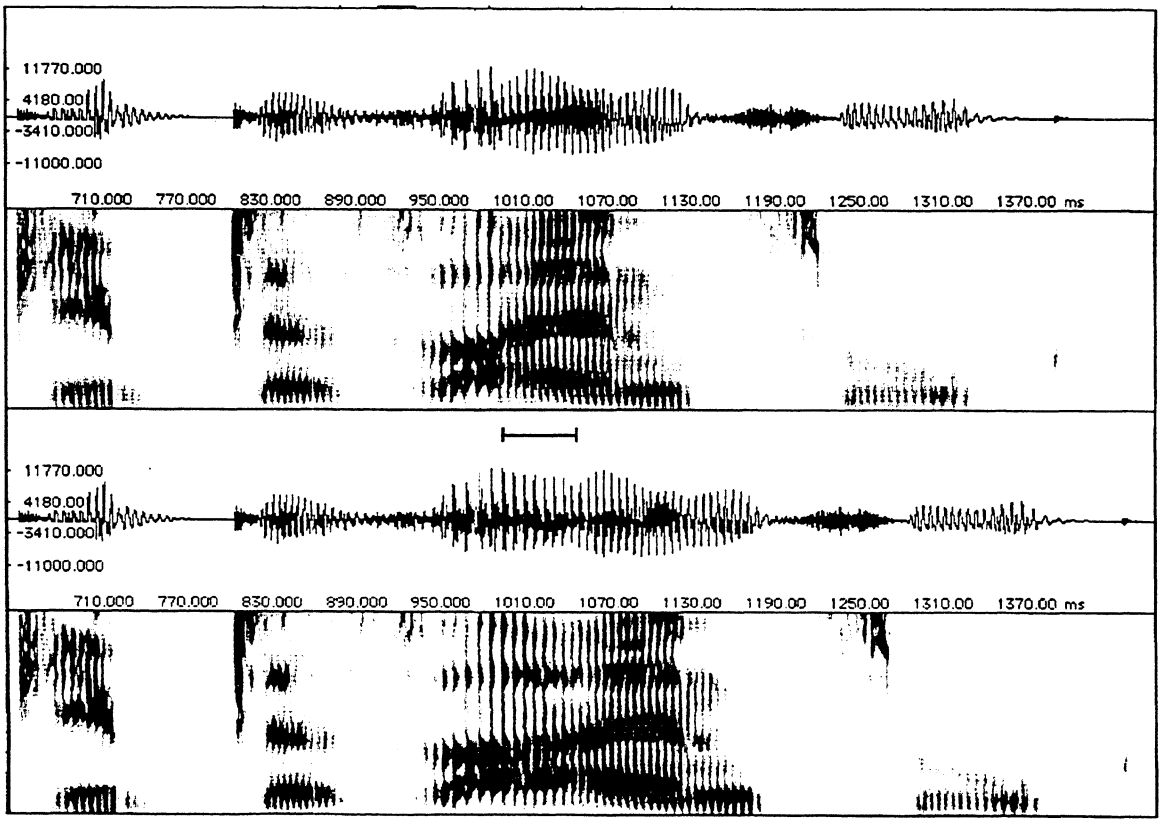


Fig. 6: Audio signals and sonagrams of the manipulated (top) and the original (bottom) sentence part "Gibt es einen Zug ...?" (s063 of speaker hpt): In the manipulated utterance the marked [a] segment of 49 ms was cut.

The results of the present pilot investigation clearly show that much can be learned about the articulatory organization in fluent speech by a thorough acoustic analysis (accompanied by perceptual evaluation) of larger data bases of connected/spontaneous speech as are only recently available for German through the PHONDAT project.

REFERENCES

IPDS (1995), CD-ROM#1: The Kiel Corpus of Read Speech, vol. I. Kiel.

Kohler, K.J. (1990), Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle, W.J. & Marchal, A. (eds.), Speech Production and Speech Modelling. Dordrecht, 69-92.

Kohler, K.J. (1992), Gestural reorganization in connected speech: a functional viewpoint on 'articulatory phonology'. Phonetica 49, 205-211.

Kohler, K.J. (1994), Lexica of the Kiel PHONDAT Corpus Read Speech, Kiel [= Arbeitsberichte Institut für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK) 27 & 28].

Kohler, K.J. (1996), Articulatory reduction in German spontaneous speech. In: Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modelling & 4th Speech Production Seminar. Autrans, 1-4.

Kröger, B.J. (1993), A gestural production model and its application to reduction in German. Phonetica 50, 213-233.

Pompino-Marschall, B. (1992), PHONDAT. Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank für gesprochenes Deutsch. In: Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 30, 99-128.

Tillmann, H.G. (1995), Kleine und Große Phonetik. Phonetica 52, 144-159.

Wells, J.; Barry, W.; Grice, M.; Fourcin, A. & Gibbon, D. (1992), Stage report SEn.3 Standard computer compatible transcription (SAM-UCL-037). In: ESPRIT project 2589 (SAM) Multilingual Speech Input/ Output Assessment, Methodology and Standardisation. Final Report.

Wesenick, M.-B. (1994), Entwurf eines überspezifizierenden Regelsystems der Aussprache des Deutschen als Basis für empirische Untersuchungen (M.A. thesis, München).

APPENDIX

Orthographic text and canonical pronunciation (in SAMPA[8]) of the selected sentences of the Kiel corpus

e068:
Wann geht bitte *ein* Zug nach Hamburg mit möglichst wenig, ja, Bahnhöfen?
v a n+ g 'e: t b 'I t @ Q aI n+ t s 'u: k n a: x+ h 'a m b U 6 k m I t+ m '2: k l I C s t
v 'e: n I C , j a:+ , b 'a: n #h "2: f @ n ?

e083:
Was ist die billigste Möglichkeit für *ein* Bahnticket von hier nach Kiel?
v a s+ Q I s t+ d i:+ b 'I l I C s t @ m '2: k l I C k aI t f y: 6+ Q aI n+ b 'a: n #t "I k I t
f O n+ h 'i: 6 n a: x+ k 'i: l ?

e097:
Wann fährt bitte *ein* Zug von Regensburg nach Dortmund am Montag, am Montagmorgen?
v a n+ f 'E: 6 t b 'I t @ Q aI n+ t s 'u: k f O n+ r 'e: g @ n s b U 6 k n a: x+
d 'O 6 t m U n t Q a m+ m 'o: n t a: k , Q a m+ m 'o: n t a: k m 'O 6 g @ n ?

e076:
Gibt es *einen* Zug, der in Köln länger hält und nach Dortmund weiterfährt?
g 'i: p t Q E s+ Q aI n @ n+ t s 'u: k , d e: 6+ Q I n+ k '9 l n l 'E N 6 h 'E l t Q U n t+
n a: x+ d 'O 6 t m U n t v 'aI t 6 #f "E: 6 t ?

s055:
Gibt es *einen* Nachtzug, der am dreiundzwanzigsten Zwölften morgens in Ulm ankommt?
g 'i: p t Q E s+ Q aI n @ n+ n 'a x t #t s "u: k , d e: 6+ Q a m+
d r 'aI U n t #t s v "a n t s I C s t @ n t s v '9 l f t @ n m 'O 6 g @ n s Q I n+ Q 'U l m
Q 'a n #k "O m t ?

s063:
Gibt es *einen* Zug, der am dreiundzwanzigsten Zwölften um zirka zwanzig Uhr in Oldenburg ist?
g 'i: p t Q E s+ Q aI n @ n+ t s 'u: k , d e: 6+ Q a m+ d r 'aI U n t #t s v "a n t s I C s t @ n
t s v '9 l f t @ n Q U m+ t s 'I 6 k a: t s v 'a n t s I C Q 'u: 6 Q I n+ Q 'O l d @ n b U 6 k Q I s t+ ?

[8] In the PHONDAT version of SAMPA the symbol + is added as a marker to the last segment of unstressed function words and the symbol # denotes compound boundaries.