Some observations on *'ein'* and *'einen'*

Bernd Pompino-Marschall and Peter M. Janker
Zentrum für Allgemeine Sprachwissenschaft, Belin

Introduction

The paper addresses the question of what distinguishes the two word forms *'ein'* and *'einen'*, i.e. the German indefinite article nominative singular (mask., neutr.) vs. accusative singular (mask.) when produced in connected speech.

Generally, it is assumed that [ən] endings in German by [ə] deletion reduce to syllabic [n̩] in casual pronunciation. This for example seems to be litterally true for German content word forms ending in [nən]. The material analysed in preparation of the planned new edition of the WdA in about 80% of the cases showed [ə] deletion and more than 50% of these reduced forms showed clear indications of two separable [n] sounds, i.e. differences in the energy or f_0 contour (Siegrun Lemke, personal communication). For the function word contrast between *'ein'* and *'einen'* however, inspection of the Kiel corpus of read speech in almost no cases of reduced *'einen'* revealed reflexes of two [n] sounds in the energy or f_0 contour even in cases where two [n] segments were labelled.[1] The distinction of these word form tokens from *'ein'* therefore seems to rely on the different length of the [n] sounds, i.e. [aɪnː] vs. [aɪn] (as against assumed [aɪnn̩] vs. [aɪn]).

1 The perception of *'ein'* and *'einen'*

In order to test this hypothesis that hearers are capable of differentiate between both word forms cued by [n] duration only, we constructed a couple of listening experiments with manipulated naturally produced acoustic speech material. A prototypical token of reduced 'einen' was cut from the utterance (RTDS046) with the help of the Signalyze 3.12 software for Apple Macintosh. It consisted of a glottalized [a] segment of 68 ms, a modal diphthongal segment of 65 ms and a [n] segment of 49 ms duration (cf. figure 1).

Procedure

The duration of the word final [n] was modified the following way: (1) two step shortening of 10 ms each by cutting the fifth and sixth or the fourth to seventh pitch period counted from the end of the acoustic signal and (2) two step lengthening by doubling the pitch periods five and six or four to seven. Since it was expected that the duration based distinction is dependent on speech rate, the duration of the diphthongal segment was also manipulated in equal steps of 10 ms by cutting/doubling two or four pitch periods. The pitch periods for this manipulation were chosen in a way to keep the formant transition of the diphthong intact (cf. figure 1). This resulted in 25 stimuli (5 [n] durations * 5 [aɪ] durations: -20, -10ms, original, +10, +20 ms each) that were presented with and without the initial glottalisation five times in quasi randomized order (resulting in 250 items) to the subjects. The subjects listened to the stimuli via headphones at a comfortable listening level in a quiet room. They marked their identification responses (*'ein'* or *'einen'*) on prepared answer sheets.

---

[1] Here, some slight spectral changes at the supposed segment boundary may be detectable.

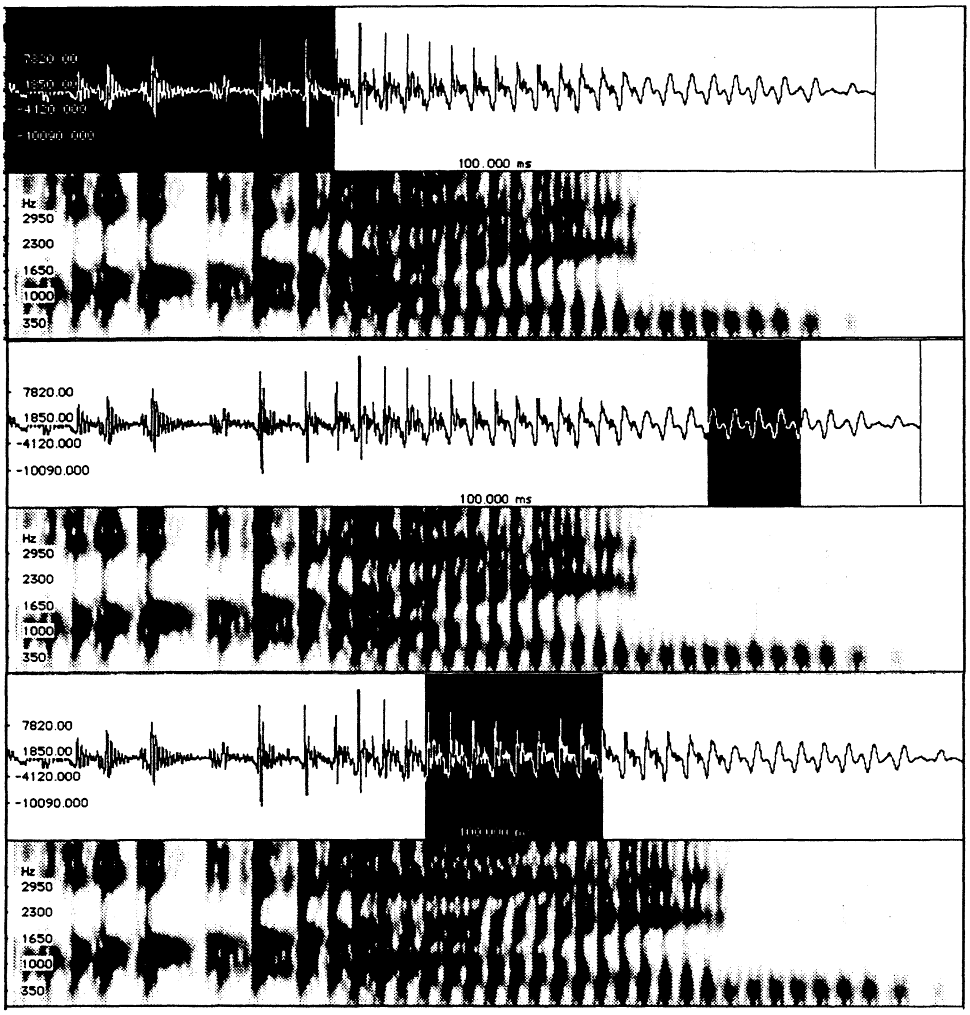

Fig. 1: Oscillograms and sonagrams of selected stimuli: The original signal with the initial glottalisation marked (top); first step of [n] lengthening with the two doubled pitch periods marked (mid); second step of [aɪ] lengthening with the four duplicated pitch periods marked (bottom).

Results

The results of this listening test averaged over 17 subjects are depicted in figure 2. The raw data was subjected to an analysis of variance with the factors [n] duration, [aɪ] duration and glottalisation and the number of *'einen'* responses as the dependent variable.

Analyses of variance revealed a highly significant ($p << .001$) effect of [n] duration, presence of glottalization as well as a significant ($p < .01$) interaction of both effects and a marginal effect ($p < .05$) of vowel length (distinguishing between vowel length 2 and 5; cf. fig. 1). Glottalized items were generally more often identified as *'einen'*. For the glottalized as well as for the non-glottalized items only a highly significant ($p << .001$) effect of [n] duration remained: The longer the duration of the nasal segment the more *'einen'* responses. For post hoc Scheffe comparisons of pairs for glottalized items only pair 1/2 and 4/5 and for non-glottalized items additionally pair 2/3 failed to reach significance.

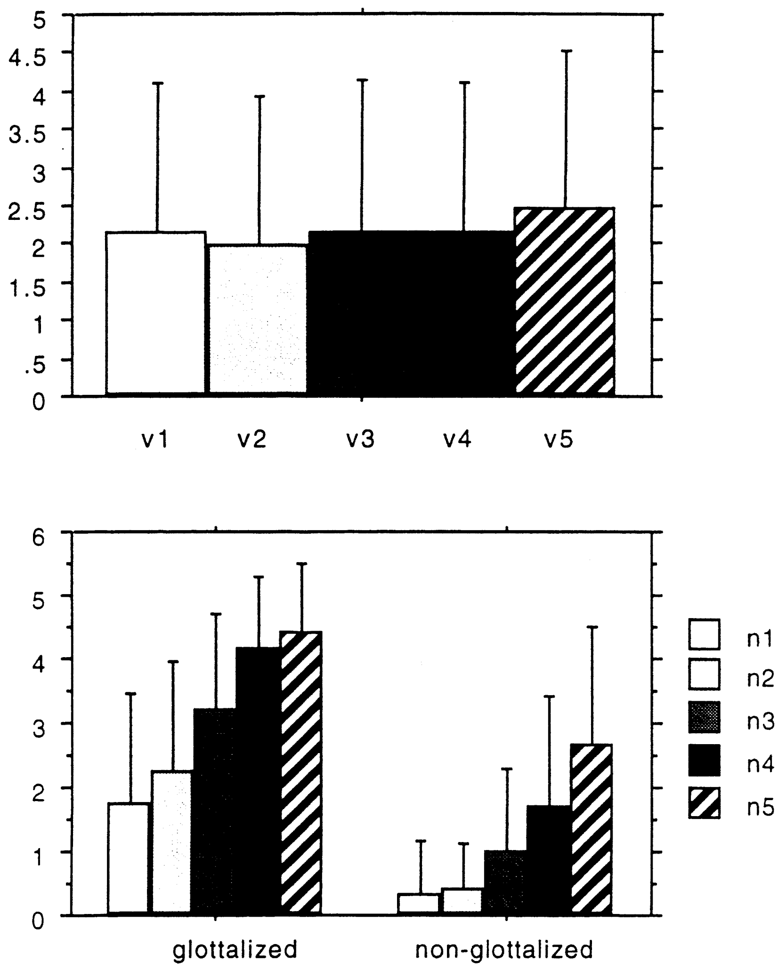

Fig. 2: Results of the listening experiment: Mean *'einen'* responses to stimuli with diphthongal segments of different length (top); interaction of [n] duration and presence/absence of glottalisation (bottom; error bars represent one s.d.).

For the perception of socalled syllabic [n] in our material the duration of the nasal segment - independent of the duration of the modally voiced vocalic portion but reinforced by the presence of glottalization (i.e. a longer vocalic portion) and therefore independent of speech rate - seems to be the only reliable segmental cue. We will take up the effect of glottalization again in the general discussion.

2 The production of *'ein'* and *'einen'*

Recording procedure

Figure 3 depicts the general experimental setup. Tongue movements were monitored by means of electromagnetic articulography (AG100 Carstens Medizinelektronik, Göttingen, Germany). This method involves the use of three transmitter coils (mounted on a helmet) to generate an alternating magnetic field at three different frequencies. The field strength detected by sensor coils mounted on the articulators is roughly inversely proportional to the cube of the distance between sensor and transmitter (see Perkell et al. 1992, 1993; Schönle 1988 for background to electromagnetic transduction systems). The raw distance signals are then converted by software to x-y coordinates in the midsagittal plane. In order to guarantee the quality of the articulatory data, additional procedures were implemented allowing more accurate calibration and better detection of unreliable data (see Hoole 1993 for details).

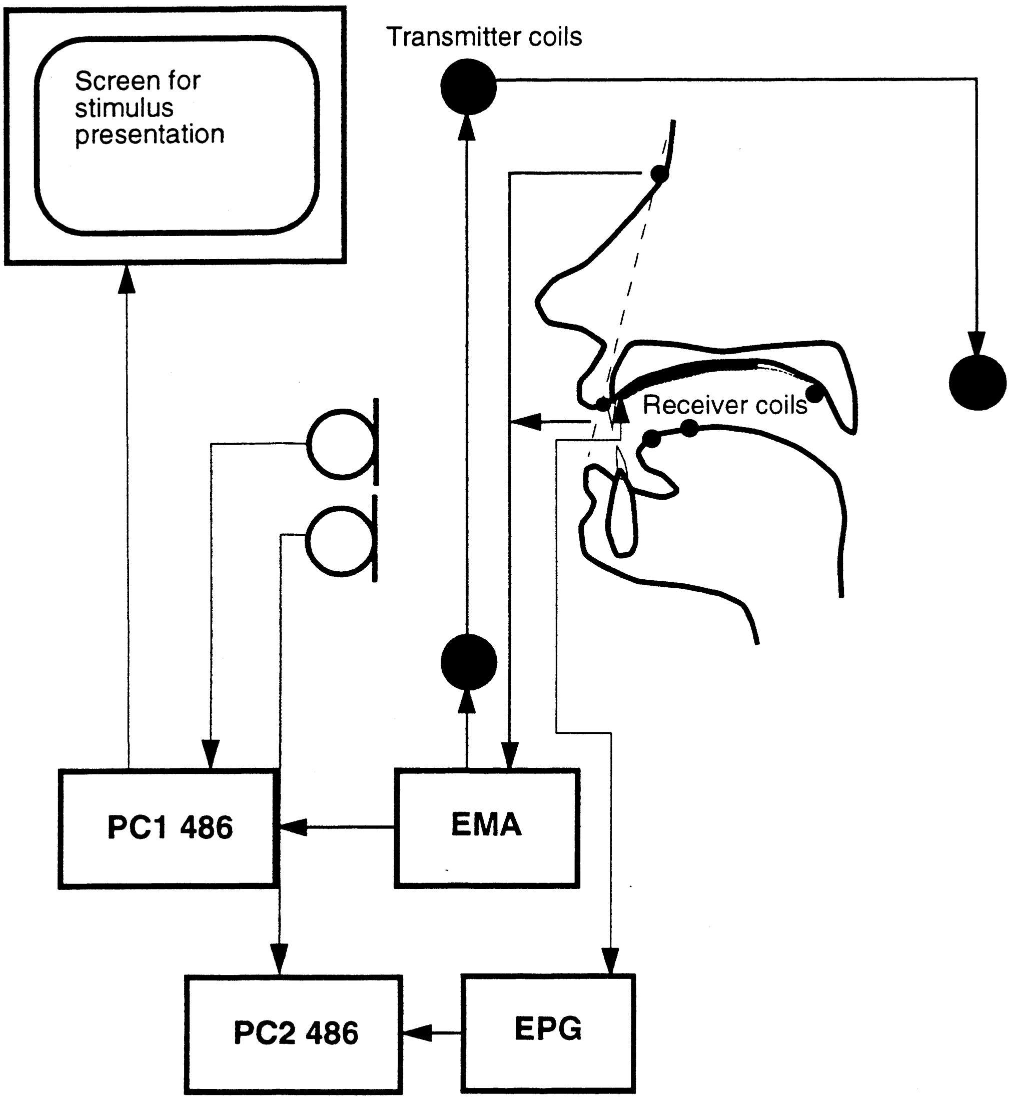

Fig. 3: Experimental setup and placement of receiver coils (bottom right: front view of the subject with tongue streched out to demonstrate the placement of the receiver coils).

Details of the sensor positions are as follows: Two transducers were mounted on the midline of the tongue at about 1 and 5 cm from the tongue tip (henceforth TB - tongue blade - and TD - tongue dorsum coil, respectively). The third coil was mounted on a strip of elastic foil glued to the back of the artificial EPG palate touching against the back part of the velum when the palate is inserted (henceforth V; cf. figure 4). Two reference coils were attached to the upper incisors and the bridge of the nose to correct for head movements.

The modified recording software (Hoole 1993) stored the movement data of the five receiver coils (recorded at 400 Hz) together with the information of the instantaneous tilt and the synchronous audio signal (16 bit, 16 kHz) in compressed form on a PC.

Besides the articulatory data at the end of the test session a tracing of the hard palate of the subject was made by using a sensor attached to the finger of one of the investigators.

The raw data were preprocessed to (1) correct for the remaining measurement error[2], (2) rotate to the vertical axis defined by the positions of the coils at the bridge of the nose and the upper incisors, (3) decompress the audio file, and (4) splitting the tilt data from the position data.

Tongue palate contacts were measured by means of the Reading EPG3 system with an artifical palate with 62 electrodes (seven rows of eight plus front row of six electrodes) every ten ms parallel to the audio recording with another PC (cf. figure 3 & 4).

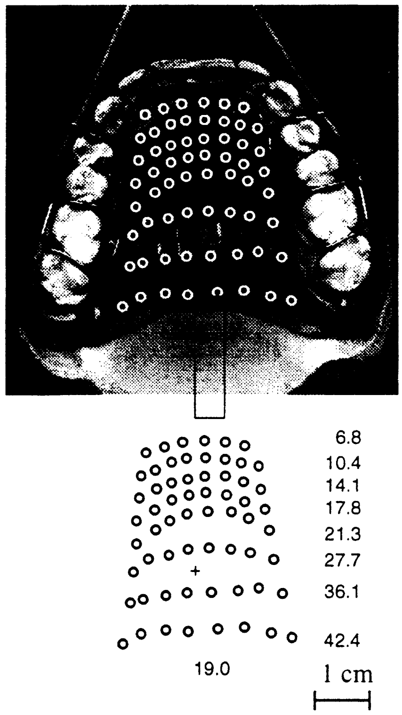

Fig. 4: EPG palate (above) and scheme of electrode placement (below) for subject JDR (the rectangle marks the strip of foil glued to the palate to carry the velar EMA coil; mean distances [in mm] of electrode rows from the inner edge of the upper incisors is given right, the higest point from bite plane [distance in mm given below] is marked by a cross).

Material

The male native German subject (JDR) read parallel constructed sentences with 'ein' and 'einen' in randomized order five times each in three different recording blocks. First in his normal pronunciation, the second time more carefully and then again more quickly and casually. The sentences had the form of

'Es fuhr ein Audi nach Augsburg' ('An Audi was going to Augsburg')

'Er fuhr einen Audi nach Augsburg' (He drove an Audi to Augsburg)

(parallel sentences contained *'Kombi'* ('utility car') / *'Cottbus'*, *'Traktor'* ('tractor') / *'Trabach'*, *'Volvo'* / *'Wolfsburg'*, *'Mazda'* / *'Monza'*).

---

[2] By using the computed error during calibration.

Analysis procedures

Durational measurements in the acoustical signal were conducted with the Signalyze software for Apple Macintosh. In the acoustic signal the following durations/time points were determined manually under auditory and visual (especially sonagraphic) feedback: the beginning of the initial glottalisation, the beginning of the diphthong, the beginning and the end of the first [n] segment as well as (when applicable) the beginning of the [ə] segement and/or the beginning and the end of the second [n] segment.

The EPG data was analyzed (1) with respect to the duration of linguopalatal contact during the production of the nasals and (2) with respect to the position of the centre of gravity of the area of linguopalatal contact averaged over all frames that show at least one electrode row of total closure. As in Gibbon et al. (1993) the centre of gravity was computed over the contacts of the central four midsagittal electrodes of the anterior fo rows of electrodes. In case of [ə] elision in items of *'einen'* the centre of gravity was calculated separately for the first and the second half of the total contact duration. For cases of assimilatory changing of the nasal place of articulation (in items of *'Kombi'* / *'Cottbus'*) also the duration of closure overlap at the alveolar and the velar place of articulation was measured.

The EMA data was analyzed with respect to the alveolar closing/opening behaviour of the TB coil and the velar lowering/raising behaviour of the V coil. Minima within the tangential velecity function were used as starting points for gestural analysis. On- and offsets of movement were defined as 20% points within the total flesh point displacement function. Besides the maximal velocity the duration as well as the fleshpoint position at the beginning and the end of the gesture and the intergestural timing (e.g. velar hold - the interval between the end of velar lowering and the beginning of velar raising) were determined.

The data were subjected to analyses of variance with speaking style (careful, normal, fast), produced word form (*'ein'*, *'einen'*) and following consonant (zero, labial, dental, alveolar, velar) as independent variables.

2 Results

The results of the durational measurements on the audio signal are reported in table 1.

Table 1: Acoustical segment durations

word	style	~	aɪ	n	ə	n
ein	careful	23.4 (15.75)	55.9 (21.87)	83.0 (21.78)		
	normal	34.8 (16.72)	83.5 (15.61)	98.6 (29.66)		
	fast	.4 (2.10)	51.3 (11.49)	58.2 (14.43)		
einen	careful	24.6 (15.59)	47.5 (15.25)	62.2 (70.63)	73.8 (19.36)	79.1 (25.03)
	normal	27.4 (13.98)	62.0 (15.48)	248.1 (41.08)	•	•
	fast	1.0 (3.56)	61.6 (11.89)	107.0 (24.51)	•	•

Analyses of variance revealed a highly significant ($p << .001$) effect of speaking style on the duration of the word-initial glottalization due to the fact that it is nearly totally absent (in 94% of the cases) in the fast productions of our speaker. The duration of the modal diphthong as well as the vocalic part as a whole showed a higly significant ($p << .001$) effect of speaking style, a significant ($p < .05$) effect of word form and also a higly significant ($p << .001$) interaction of both effects. These vocalic parts are significantly ($p < .01$) shorter (-21.6/-29.0 ms) in the normal but also significantly ($p < .01$) longer (10.4/11.0 ms) in the fast productions of the word form *'einen'* than in *'ein'*. The (first) nasal segment showed highly significant ($p << .001$) influences of speaking style for both word forms and for *'ein'* also of consonantal context ($p << .001$) as well as an interaction of both effects ($p < .01$). Simple effects show up as different significant differences (none in alveolar context) within the general ranking 'fast < careful < normal' and significant differences in the ranking between alveolar and labiodental (mean difference 43.2 ms).

A preliminary analysis of the EPG and EMA data showed results that in some cases seem to contradict the acoustical measurements. So, for example, in seven cases (i.e. 28%) of the normal *'einen'* productions there was no perfect alveolar closing contact resulting in different segmental durations when measured by EPG. For these items, on the other hand, the EMA data didn't show significant differences in the amount of vertical movement of the TB coil.

The EMA analysis of the *'einen'* utterances revealed that despite an always present slight tongue tip lowering of about 4 mm (in contrast to the elevation of 11.6 mm for producing the alveolar closure of the [n]) for the [ə] production in careful style, the velum does not show corresponding closing movements but remains open during the vowel production. This velar lowering is, on the other hand, with 1.9 mm significantly less ($p << .001$) than for the normal and fast productions (3.8 mm) and starts on the average 29.3 ms later than the tongue blade movement in significant contrast ($p << .001$) to the normal and fast productions, where the velar gesture precedes the tongue blade movement by about 96.3 ms. The tongue tip lowering for [ə] was only once observed in normal speaking style, never at fast speech rate.

As to be expected, the variation of the consonantal context affected the position of the alveolar contact for the nasal: The position of the alveolar closure as determined by the center of gravity of the EPG contact pattern showed significant influences of the speaking style, the consonantal context as well as their interaction for the whole nasal segment in *'ein'* ($p << .001$; $p < .05$; $p < .05$) as well as the second (or the second half of the) nasal segment in *'einen'* ($p << .001$; $p < .01$; $p << .001$). Split by word form and speaking style the simple effects were as follows: The contact in the nasal segment of *'ein'* in the velar context is 0.18 rows more backwards than in the zero and alveolar context in careful pronunciations ($p < .01$), 0.30 and 0.32 rows more backwards than in all other contexts in normal and fast productions (both $p << .001$); for the careful *'einen'* utterances there was an only marginal ($p < .05$) effect (zero context 0.09 more backwards than alveolar context), but for the normal and fast productions the velar context again showed more backward contacts in the velar context (0.28 rows in contrast to all other contexts for the normal productions and 0.13 rows in contrast to the labial and labiodental context; both $p < .001$).

The nasal productions in the velar context also showed an assimilatory overlapping (of about 55 ms) of alveolar and velar contacts that is marginally significant ($p < .05$) dependent on speaking style: With 64 ms this overlap is 19 ms longer in fast speech rate than in normal productions.

4 General discussion

The experiments described above demonstrate the large variability of word form realizations in this quite simple example of *'ein'* vs. *'einen'*.

For the hearer, the most reliable cue for bisyllabic *'einen'* seems to be the pure length of the nasal segment. But there is also an effect of the presence of an initial glottalization.[3] These more frequent 'einen' responses to glottalized items may be due to the fact that this glottalization is perceived as an event of its own, not belonging to the following syllable, i.e. resulting in a bisyllabic percept. To cite an example of a natural realization of *'einen'* already described in Pompino-Marschall (1996) there seems to be the possibility of signalling the bisyllabicity of *'einen'* not by a syllabic [n̩] but - anticipated - as a bisyllabic vocalic segment: The original utterance shown in figure 5 is unambiguously perceived as *"Gibt es einen Zug ...?"* but when the steady-state [a] portion of 49 ms duration is cut - as shown in figure 5 (top) - the ungrammatical utterance * *"Gibt es ein Zug ...?"* is heard.

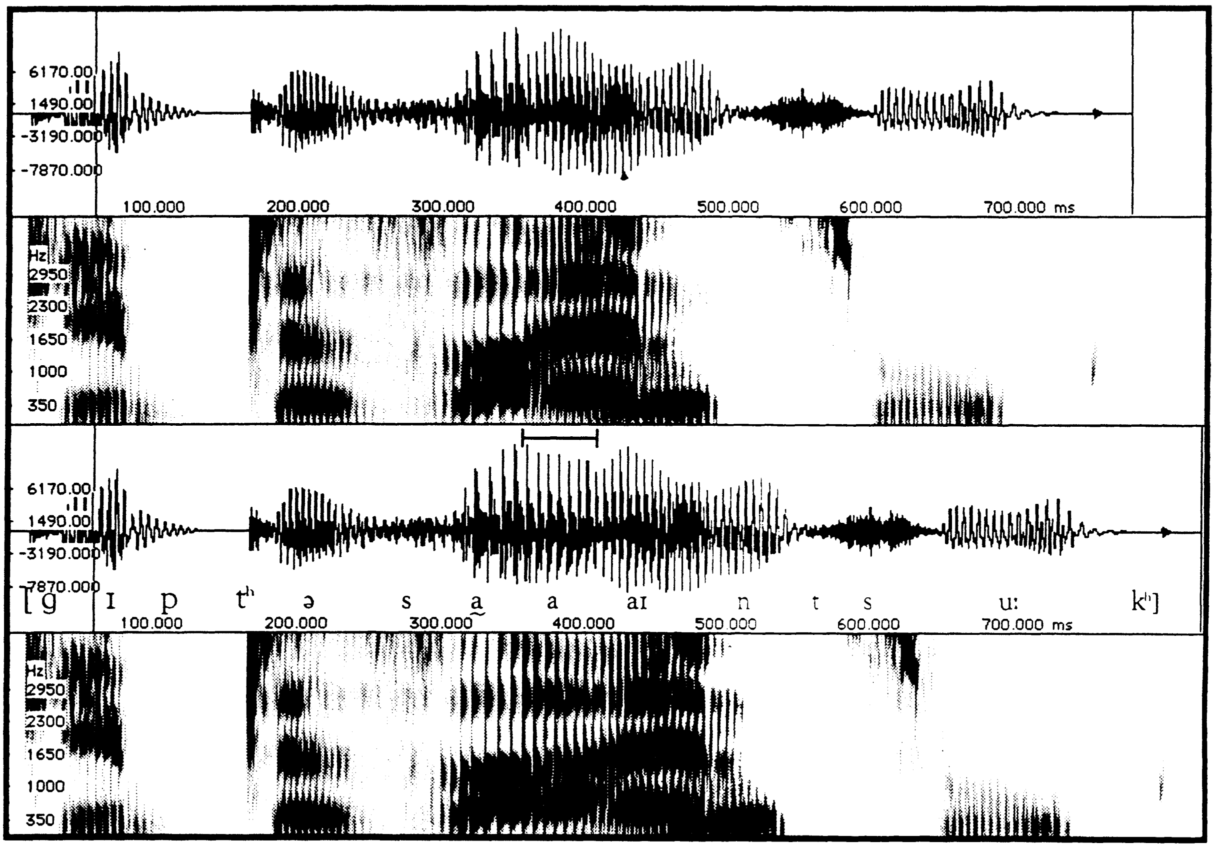

Fig. : Audio signals and sonagrams of the manipulated (top) and the original (bottom) utterance "Gibt es einen Zug ...?" (HPTS063): In the manipulated utterance the marked [a] segment of 49 ms was cut.

In order to test this possibility further, some more test tapes were prepared where the preceding word *'noch'* of the original utterance was pasted before the manipulated items of *'einen'* of the listening tests described above. This procedure yielded stimuli in which the glottalization is perceived as an integral part of the following syllable. To compensate for a perceptually resulting speech rate accelleration in the stimuli with deleted glottalizations in these items also a silent

---

[3] N.B. in opposite direction of a compensatory shortening of the vocalic part of the utterance.

interval of the duration of the glottalized segment was inserted. To test the influence of the diphthong - besides its duration - on the perceived dichotomy another test with synthetic material was constructed where the amount of the F1/F2-transition from [a] to [ɪ] was varied systematically.

The articulatory investigation, on the other hand, showed that it is by no means a simple segmental process that underlies the different pronunciation variants of *'einen'*. There is not only a simple deletion or reduction of single gestures but also a complex restructuring of the inter-articulator coordination in timing as well as in amount.

References

Gibbon, F.; Hardcastle, W. & Nicolaidis, K. (1993), Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. Language and Speech 36, 261-277.

Hoole, P. (1993), Instrumentelle Untersuchungen in der artikulatorischen Phonetik: Überlegungen zu ihrem Stellenwert als Grundlage für Entwicklung und Einsatz eines Systems zur Analyse der räumlichen und zeitlichen Strukturierung von Sprechbewegungen. München [phil.Diss.].

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

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.

Perkell, J.S.; Cohen, M.; Svirsky, M.A.; Matthies, M.L.; Garabieta, I. & Jackson, M.T.T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. Journal of the Acoustical Society of America 92, 3078-3096.

Perkell, J.S.; Svirsky, M.A.; Matthies, M.L.& Manzella, J. (1993), On the use of electromagnetic midsagittal articulometer (EMMA) systems. In: Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 31, 29-42.

Pompino-Marschall, B. (1996), Articulatory reduction in fluent speech: A pilot study on syllabic [n] in Standard German. ZAS Papers in Linguistics (ZASPIL) 7, 151-162.

Schönle, P. (1988), Elektromagnetische Artikulographie. Berlin.

Aknowledgement

Work supported by German Research Concil (DFG) Grant Po 334/2 to the first author.