PRODUCTION AND PERCEPTION OF SYLLABIC [n] IN GERMAN

Bernd Pompino-Marschall and Peter M. Janker
ZAS - Research Centre for General Linguistics, Berlin, Germany

ABSTRACT

Production and perception of syllabic [n̩] was studied using different inflectional forms of the German indefinite article, i.e. *ein* (nom. sing. masc.) with simple [n] vs bisyllabic *einen* (acc. sing. masc.) that is realized in a reduced manner with a so called syllabic [n̩]. Inspection of a larger corpus of spoken Standard German revealed that the most frequent pronunciation of *einen* is the allegro form with a lengthened [nː], normally without any indication of an extra syllable. Production was analyzed in a combined electropalatographical (EPG)/electromagnetic-articulographical (EMA) study. The articulatory analysis revealed quite complex gestural reductions for bisyllabic *einen*. Identification experiments with manipulated material showed that the distinction between *ein* and *einen* is dependent on the duration of the nasal alone, independent of local articulatory tempo but not of global speech rate.

1. INTRODUCTION

Generally, it is assumed that in German by [ə] deletion (cf. [2]) [ən] endings reduce to syllabic [n̩] in casual pronunciation. For example, this seems to be true for German content words ending in [nən]. For the function word contrast between *ein* and *einen* however, a preliminary inspection of the 'Kiel Corpus of Read Speech' [2] showed that in almost all cases of reduced *einen* there are no indications of two [n] sounds in the energy or f0 contour even in cases where two [n] segments were labelled. 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]).

2. PRODUCTION OF *ein* AND *einen*

The production of the different realization of the word forms *ein* and *einen* was studied by an analysis of a larger database of read and spontaneous speech [2-5] and in a controlled production experiment.

2.1. Data Base Analysis

The results of the data base analysis[1] is depicted in figure 1. In the 233 tokens of *einen* the ending [ən] is articulated in only 11.2% (there are 41.3% [ə] elisions and 46.8% elisions of both [ə] and [n]). When [ə] is lost, the segment boundary between the adjacent nasal segments in 77.1% of these cases was marked as being not secure.

Besides the variation in pronunciation a large variation of segmental durations within the single types is clearly visible in figure 1. This variation surely is partly due to effects of speech rate.

To test the influence of speech style/rate on the variation of pronunciation and on segmental duration the following controlled production experiment was run.

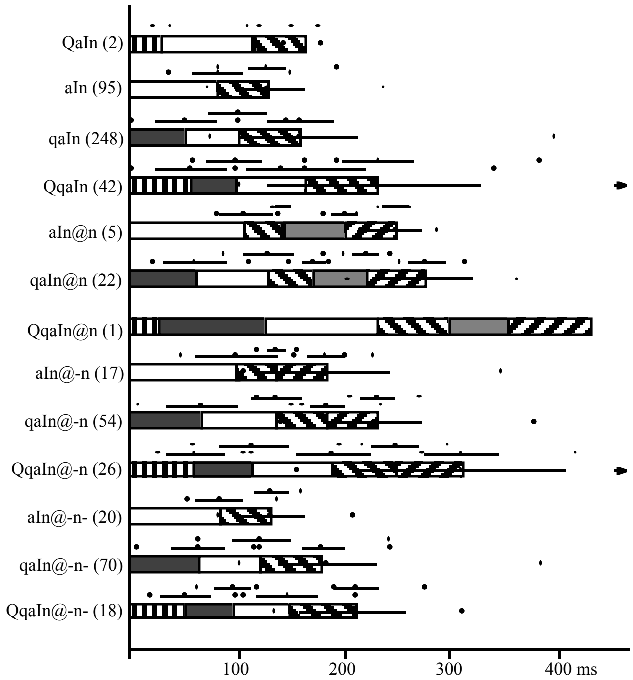

Figure 1. Mean segmental durations of the different realizations of the word forms *ein* and *einen* (glottal stop [Q][2]: vertical hatch; glottalization [q]: dark grey; modal diphthong [aI]: white; first [n]: falling hatch; Schwa [@]: light grey; second nasal: rising hatch; the number after the transcriptions represent the the number of tokens; lines above the bars represent ± 1sd of the single segment durations while points mark the extreme values; lines and points at the center of the bars represent the durational variation of the whole word utterances, arrows indicating extreme values outside the range of the scale.

2.2. Production Experiment

2.2.1. Procedure. Three male native German subjects (JDR, PMJ, DFI) read parallel constructed sentences with *ein* and *einen* in randomized order five times each under three different speech rate conditions: normal tempo, more carefully and fast. 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'[3]

Besides the audio signal articulatory data were recorded by midsagittal electromagnetic articulography (EMA, Carstens AG 100, 5 channels) and by electropalatography (EPG, Reading system 3.0 with 62 electrodes in 8 rows) in parallel.

The EMA (cf. [7]) sensor coils were mounted on the tongue at about 1 and 5 cm distance from the tip (henceforth TB - tongue blade - and TD - tongue dorsum coil, respectively). A 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).[4] 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 velocity function were used as measuring points for gestural analysis.

The EPG data was analyzed (i) with respect to the duration of linguopalatal contact during the production of the nasals and (ii) with respect to the position of the center of gravity of the area of linguopalatal contact averaged over all frames that show at least one electrode row of total closure. The center of gravity was computed over the contacts of the central four midsagittal electrodes of the anterior four rows of electrodes (cf. [1]). In case of [ə] elision in items of *einen* the center 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 data were subjected to separate analyses of variance for each speaker with speech style (careful, normal, fast), word form produced (*ein*, *einen*) and following consonant (zero, labial, dental, alveolar, velar) as independent variables.

2.2.2. Results. The recorded data were fully analyzable for subject JDR only. PMJ's data showed a general velar lowering with almost fully absent active velar movement[5] and DFI's velar measurements are not reliable due to technical reasons. The results are therefor first described for subject JDR in more detail.

Analyses of acoustical segment durations revealed a highly significant ($p << .001$) effect of speech 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 'alveolar < labiodental' (mean difference 43.2 ms).

The analysis of the articulatory data showed results that at first sight seem to contradict the acoustic 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* ($p << .001$; $p < .01$; $p << .001$). 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.

Besides a generally slower tempo in careful style PMJ's productions are characterized by more marked glottalizations that are, on the other hand, totally absent in his generally more reduced fast productions. His *einen* utterances in careful and normal style show resistance to schwa elision (only two exceptions). The fast productions of PMJ, in contrast, exhibit a total absence of [ə] but only marginally compensatory lengthening of the nasal segment.

Compared to JDR an overlap between alveolar and velar closure can be seen quite rarely in the EPG data but in normal and fast productions of *ein* (not in the second nasal of *einen*) in 60/40% there is a total assimilation, i.e. no alveolar contact at all. In alveolar context (*Traktor*) a fronting of articulation of the preceding nasal segment is observable (e.g. in normal productions of *ein* gravity is 1.97 vs 2.51 in non-alveolar context).

Subject DFI's data exhibit no influence of rate on glottalization but parallel to PMJ only slight temporal compensation for schwa elision in nasal duration. His temporal variation, on the other hand, resembles that of JDR.

DFI's EPG data show less influence of consonantal context with the exception that there is an almost 100% total assimilation of place of nasal articulation in velar context.

These quite different data 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 speaker dependent complex restructuring of the interarticulator coordination in timing as well as in amount.

3. PERCEPTION OF SYLLABIC [n̩]

3.1. The Relevance of Segmental Duration

In order to test the hypothesis that hearers are capable of differentiating between *ein* and *einen* cued by [n] duration alone, we constructed a couple of listening experiments with manipulated naturally produced acoustic speech material.

3.1.1. Procedure. A prototypical token of reduced *einen* was cut from the utterance with the help of the Signalyze 3.12 software for Apple Macintosh. The original utterance consisted of a glottalized [a̰] segment of 68 ms, a diphthong of 65 ms and an [n̩] of 49 ms duration produced at a fundamental frequency of about 200 Hz by a female speaker. The duration of the diphthong and the nasal were varied in five steps of approximately 10 ms by doubling/cutting pairs of individual pitch periods. A second set of 25 stimuli was produced by cutting the initial glottalized segment. 17 native speakers of German served as subjects in the identification test consisting of five repetitions in randomized order. They had to decide whether *ein* or *einen* was uttered.

3.1.2. 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. 2). 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.

3.2. The Influence of Speech Rate

3.2.1. Procedure. Since the glottalization in the test stimuli may be perceived as not belonging to the same syllable as the rest of the test word, we constructed another test, where the preceding word *noch* ‘another’ of the original utterance was pasted before the manipulated items of *einen*. 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 a second set of these items a silent interval of the duration of the glottalized segment was inserted instead.

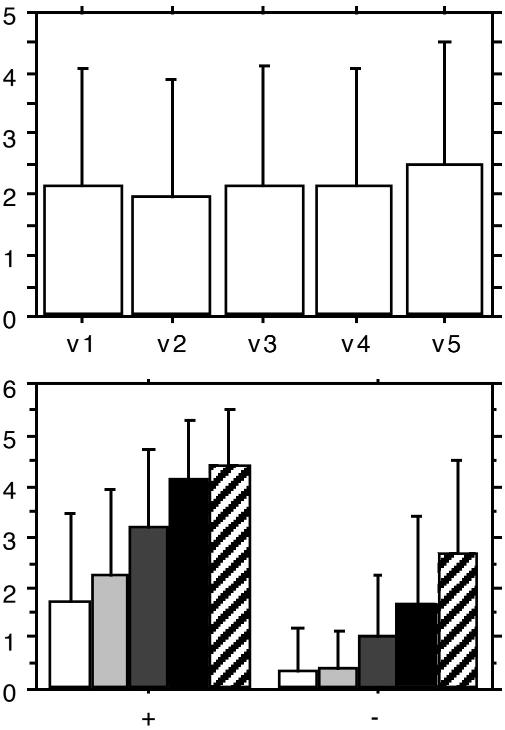

Figure 2. Mean *einen* responses as a function of vowel length (above; v1 (shortest) - v5) and of presence of glottalization and [n] duration (below; +: glottalized, - nonglottalized; [n] durations in ascending order; error bars represent 1sd).

3.2.2. Results. Comparing the *einen* responses in the two sets of stimuli (cf. fig. 3) the analysis of variance revealed a highly significant ($p < .001$) effect of the inserted pause as well as significant ($p < .05$) interactions with nasal duration and presence of glottalization: The items without the glottalized segment and without compensatory pause insertion are more readily perceived as *einen* and they are perceived as produced with a higher speech rate than their glottalized counterparts.

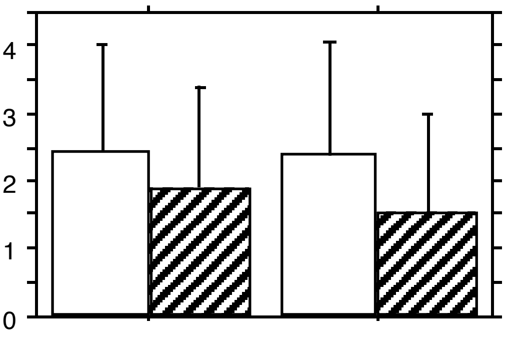

Figure 3. Comparison of the effect of the absence of glottalization (hatched bars) on mean *einen* responses when no compensatory pause is inserted (left) to items with speech rate correcting pause (right).

4. CONCLUSION

So called syllabic [n̩] in German, in contrast to normal word final [n], as studied here in the unstressed word forms of the indefinite article *ein* and (reduced) *einen* seems to be mainly characterized by a lengthened nasal segment.

The analysis of the 'Kiel Corpus' showed quite a variety of segmental variation for those word forms as well as for the single segment's durations.

Our controlled production experiment revealed that these segmental and durational changes are due to quite complex articulatory restructuring processes under the different speech rate conditions. These spatial and temporal restructuring processes are furthermore strongly speaker dependent.

The perception of syllabic [n̩], on the other hand, seems to be dependent on the duration of the nasal segment alone. Identification seems to be independent of rate of articulation but not of general speaking rate when the test items are embedded in larger utterances. This was further confirmed by a recent test that studied the influence of the tempo of a precursor phrase and of vowel duration on the perception of aspiration (VOT continuum) and 'syllabic' [n] in parallel [8]. Highly significant effects of speaking rate of the precursor phrase and of vowel length were found for both the perception of voicelessness and of 'syllabic' [n]: More voiceless responses at critical VOT durations were encountered with fast precursor and short vowel. This could be interpreted as influence of local (articulatory) tempo (cf. [9]). 'Syllabic' [n], on the other hand, was more readily perceived with fast precursor but in the long vowel context. Here, global tempo, not rate of articulation seems to be the relevant parameter.

ACKNOWLEDGMENTS

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

NOTES

1. After resegmenting with respect to glottal stop duration and duration of glottalized vocalic segments that were not consistently marked in the framework of the PhonDat project.
2. Here, the SAMPA transcription of the PhonDat project is used: x- represents deletion, -y the insertion, and x- -y a change of a segment.
3. Parallel sentences contained the items *Kombi* 'utility car'/*Cottbus*, *Traktor* 'tractor'/*Trarbach*, *Volvo*/*Wolfsburg*, *Mazda*/*Monza*.
4. Two reference coils were attached to the upper incisors and the bridge of the nose to correct for head movements.
5. This velic behavior is probably an explanation for the irregularities in the air stream measurements of this subject observed in the framework of the ACCOR project.

REFERENCES

[1] Gibbon, F., Hardcastle, W., and Nicolaidis, K. 1993. Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. *Language and Speech* 36, 261-277.
[2] IPDS 1994. *CD-ROM #1: The Kiel Corpus of Read Speech*, Vol. 1. Kiel.
[3] IPDS 1995. *CD-ROM #2: The Kiel Corpus of Spontaneous Speech*, Vol. 1. Kiel.
[4] IPDS 1996. *CD-ROM #3: The Kiel Corpus of Spontaneous Speech*, Vol. 2. Kiel.
[5] IPDS 1997. *CD-ROM #4: The Kiel Corpus of Spontaneous Speech*, Vol. 3. Kiel.
[6] 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.
[7] 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.
[8] Pompino-Marschall, B. and Janker, P.M. (forthcoming). Global and local tempo in speech.
[9] Summerfield, Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 5, 1074-1095.