

Production strategies in German spontaneous speech: definite and indefinite articles

Thomas Wesener

Institute of Phonetics and Digital Speech Processing, University of Kiel, Germany

ABSTRACT

This study examines the sentence-level phonetics of function words in spontaneous speech. Data from a corpus of spontaneous speech offer insight into the phonetic variability of the German definite and indefinite articles. Three issues are addressed: a) glottal activity at the onset of the indefinite articles; b) approximation, nasalisation and fricativisation of the initial /d/ in the definite articles; c) assimilation and deletion of final nasals in both types of articles. An attempt is made to relate these phenomena. The findings for spontaneous speech are compared to those for a corpus of read speech. The results broaden traditional accounts of the phonetics of these items and shed light on the relation of spontaneous and read speech.

1. INTRODUCTION

Function words such as the definite and indefinite article are ideal objects for the study of sentence-level phonetics since they occur frequently, and, being mostly unstressed, are prone to connected speech phenomena.

Traditional word-level accounts of the phonetics of function words prevail in pronunciation dictionaries. For example, the *Duden* pronunciation dictionary's blanket rule assigns a glottal stop to all initial vowels [1]. The *Großes Wörterbuch der deutschen Aussprache* [2] adopts a more differentiated approach, acknowledging that the glottal stop is frequently replaced by a so-called 'soft onset' in unstressed syllables not following a pause; however, it does not specify the phonetic nature of that type of onset

By contrast with [3], which examines read speech and inter alia the influence of stress on glottalisation in German, this study addresses spontaneous speech and is limited to unstressed tokens, since the number of stressed function words is small.

2. METHOD

This study is based on the *Kiel Corpus of Spontaneous Speech* as published on CDROM [4, 5, 6]; it examines 112 dialogues between 32 speakers. [7] and [8] describe the gathering and processing of the corpus by means of a calendar scenario, in which subjects communicate via headsets with the aim of arranging appointments.

The speakers (18 male, 14 female) are 20 to 60 years old, with an average of 34 years. The bulk of the speakers (27) comes from the northern part of Germany (Schleswig-Holstein, Lower Saxony, and Hamburg), and all subjects speak a northern variant of standard German which cannot easily be ascribed to a specific region.

Read speech data derive from the *Kiel Corpus of Read Speech* [9], which features 53 speakers (27 male, 26 female) different from those of the spontaneous corpus, but comparable as to age, language area, and standard variant. See [3] for a detailed description of the composition of the read corpus.

Pattern-matching tools were used to search the database for the items of interest (*cf.* [10]). This gives an overview of phonetic phenomena captured by the labelling, and is complemented by a

narrow phonetic transcription of a representative subset of the retrieved tokens, based on an auditory and instrumental investigation of the speech signal.

3. RESULTS

3.1. The indefinite articles

This section presents two topics: laryngeal activity at the onset of the vowel, and realisations of the final nasals, in particular their assimilation of place to following sounds.

3.1.1. Glottal activity at the onset of the vowel. Canonically, initial vowels are preceded by a glottal stop, represented in the transcription by the label *Q*. Where speakers do not produce a glottal stop, the label is marked deleted (*Q-*). The label *-q* stands for creaky voice. There are four possible combinations of labels: *Q* for a glottal stop only, *Q-q* for a glottal stop with laryngealisation of the following vowel, *Q-**-q* for creak without a glottal stop, and *Q-* for absence of glottal stop and creak.

Although the labels give a good general idea about glottal activity, they do not specify the whole range of phonetic phenomena that can be observed. For example, the label for a glottal stop *Q* does not always represent a 'neat' glottal stop: it may mark a leaky closure, where a slight opening at the posterior parts of the vocal folds seems to allow some air to escape, resulting in glottal friction filtered by supraglottal resonances. *-q* not only covers creak, but may also mark a harsh voice quality at the beginning of the vowel. *Q-*, standing for absence of any glottal reflex, may mark realisations where there is nevertheless an auditory impression of an interruption, which might be created by breathy voice.

	Q	Q -q	Q- -q	Q-	total
spont.	6 (2)	77 (21)	210 (57)	73 (20)	366
read	31 (5)	94 (16)	286 (49)	176 (31)	587

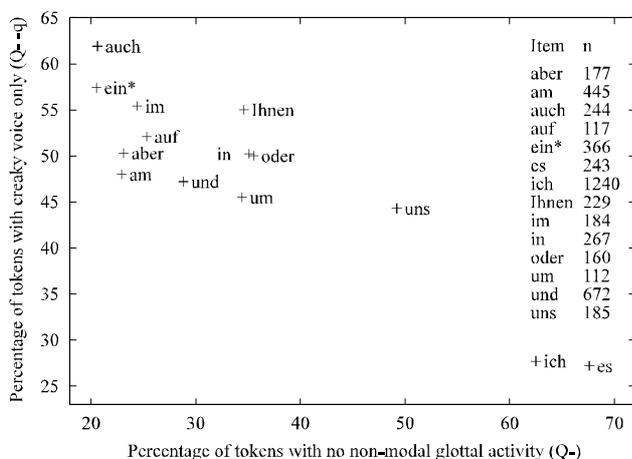
Table 1. Absolute and relative frequencies (%) of the glottal combinations in unstressed productions of the indefinite articles for both spontaneous and read speech.

Table 1 shows the frequencies of the four glottal combinations for unstressed productions of the indefinite articles in spontaneous and read speech. A chi-squared test showed the difference between the patterns for spontaneous and read speech to be significant at the 0.1% level.

In both speaking styles, the most frequent combination is creak without glottal stop (*Q-**-q*). But whereas in read speech absence of any glottal reflex (*Q-*) is the next most frequent, in spontaneous speech glottal stop with creak (*Q-q*) ranks second, closely followed by absence of any glottal reflex (*Q-*). The main difference between the two speaking styles is in cases without a glottal stop (*Q-*): creak

is more common in spontaneous than in read speech. Summing the two combinations with *-q*, creak is more common in spontaneous than in read speech (78 vs 65%). Cases with glottal stop (*Q* and *Q-q*) have roughly the same frequencies in spontaneous and read speech (23 vs 21%), so that for spontaneous speech a glottal reflex is marked in more cases. Spontaneous speech might be expected to display greater reduction than read speech, and the higher incidence of a glottal reflex in spontaneous speech seems to contradict this expectation. An explanation of this apparent contradiction may lie in the high level of disfluency phenomena in spontaneous speech: a glottal reflex is more likely after pauses, breathing, and other disfluencies typical of spontaneous speech.

Figure 1. Percentage of the two combinations without glottal stop for the unstressed indefinite articles (*ein**) and other unstressed function words with *n* > 100 for all 32 speakers (*n* indicating the



number of tokens for which the vowel has not been marked deleted).

Figure 1 relates glottal activity in the indefinite article (*ein*, *eine*, *einen*, *einem*, *eines*) to that in other unstressed function words with an initial vowel which occur at least 100 times in the corpus. For all except *uns*, *ich*, and *es*, creak (*Q-q*) is the most frequent realisation, and for all items, glottal stop on its own (*Q*) the least frequent (with its share ranging from 0.4% for *Ihnen* to 7% for *und*). Absence of a glottal reflex (*Q-*) is common; in *aber*, *auch*, *auf*, *Ihnen*, *im*, *in*, *oder*, *um*, and *und* it is the second most frequent realisation, while it is narrowly outnumbered by *Q-q* in *am* and *ein**.

Absence of a glottal reflex is more frequent than creak in *uns*, *ich*, and *es*. *ich* and *es* often appear enclitically with modal verbs, and *uns* is frequently encliticised in the pronominal sequence *wir uns*. This enclitic position differs from that of the indefinite articles and e.g. conjunctions (*oder*, *und*) or prepositions (*am*, *auf*, *im*, *in*, *um*), which occur in proclitic rather than enclitic position. Enclitic position seems to favour absence of a glottal reflex.

3.1.2. Productions of final nasals. Connected speech phenomena affect final nasals in two ways, characterised here as internal and external. The internal type occurs in *einen* and *einem*, where two nasals can mutually influence one another. *Einen* is frequently

produced with a single alveolar nasal, i.e. the nasal portion is not perceived as a geminate. This contraction is marked in 44% of cases in spontaneous speech (52 out of 119) and in 36% of cases in read speech (32 out of 89). *Einem* is most frequently realised without schwa in the second syllable, and with a bilabial instead of an alveolar nasal. The absence of the schwa is a prerequisite for the alveolar being assimilated in place to the following bilabial. The resulting nasal portion may sometimes be perceived as a geminate, and sometimes as a single sound. The existence of productions with [n]-schwa-[m] and with [nm] makes it possible heuristically to establish the line of reduction shown in note 1), of which the last two steps are the most frequent ones, occurring in 80% of cases in spontaneous speech (43 out of 54), and in 39% of cases in read speech (22 out of 57). This internal type of reduction is clearly more frequent for spontaneous than for read speech, especially for *einem*.

In the external type of phenomena, the following word influences the nasal in *ein*, *einen*, and *einem*. A typical example is *ein* being produced with a velar nasal when it precedes a velar, or with a bilabial nasal when it precedes a bilabial. The same articulations are also found for some tokens of *einen* in these contexts. *Einem* is in two cases produced with an alveolar nasal followed immediately by a velar nasal.

A further possibility is the nasalisation of the vowel preceding the canonical nasal, while the nasal itself is not realised. This occurs twice, for spontaneous as well as for read speech, before words beginning with a nasal.

In spontaneous speech, these external phenomena occur in 6% of productions of *ein*, *einen*, and *einem* (18 out of 281). They are slightly more frequent in read speech (9%, 31 out of 330); however, 18 of these 31 cases derive from readings of only 4 sentences.

3.2. The definite articles

In this section, we present phenomena connected with the initial /d/ of the definite articles and with the final nasals in *den* and *dem*.

3.2.1. The phonetics of the initial /d/. The initial alveolar plosive in the definite article is produced in three main ways. The most typical involves reduction from the canonical released plosive via an easily segmentable approximant to a minimal alveolar gesture and ultimately elision. The other directions are fricativisation following alveolar fricatives or replacement by a nasal after alveolar nasals. Table 2 shows the frequencies of the different types of production of *der*, *die*, *das*, *dem*, and *den* for spontaneous and read speech. *d-h* designates the standard case in which there is a released alveolar plosive. *d* marks cases where no release has been labelled, and mostly refers to alveolar approximants. *d-* refers to cases without a segmentable reflex of the alveolar plosive. Nasalisation is symbolised by *d-n*, and fricativisation by *d-z* or *d-s*, where *z* refers to a voiced and *s* to a voiceless alveolar fricative. Other less frequent productions are grouped together in the last column.

To turn first to spontaneous speech, the production involving a plosive and its release is clearly most common (it occurs on average in 80% of cases). The most frequent reduced form is approximation (*d* is used for 10% of the tokens).

	d-h	d	d-	d-n	d-sz	other	total

der	327 (80)	33 (8)	3 (1)	41 (10)	1 (0)	3 (1)	408
	374 (86)	29 (7)	6 (1)	12 (3)	2 (0)	13 (3)	436
die	259 (90)	15 (5)	4 (1)	8 (3)	0 (0)	1 (0)	287
	309 (88)	22 (6)	1 (0)	2 (1)	8 (2)	8 (2)	350
das	757 (74)	166 (16)	29 (3)	41 (4)	19 (2)	14 (1)	1026
	62 (73)	21 (25)	0 (0)	0 (0)	1 (1)	1 (1)	85
dem	217 (88)	8 (3)	4 (2)	18 (7)	0 (0)	1 (0)	248
	49 (94)	2 (4)	0 (0)	1 (2)	0 (0)	0 (0)	52
den	544 (84)	47 (7)	13 (2)	38 (6)	1 (0)	3 (0)	646
	103 (73)	27 (19)	3 (2)	6 (4)	1 (1)	1 (1)	140
total	2104 (80)	269 (10)	53 (2)	146 (6)	21 (1)	22 (1)	2615
	897 (84)	101 (10)	10 (1)	21 (2)	10 (1)	23 (2)	1063

Table 2. Absolute and relative frequencies (%) of different correlates of initial /d/ in unstressed *der*, *die*, *das*, *dem*, *den* for spontaneous (top row for each item) and read speech (bottom row).

An extreme case of gesture reduction is found in a production of *noch über den*, where the correlates of *den* are contained in a nasalised schwa and a drop in F1 between the preceding vowel and the schwa. This drop seems to serve as a cue to the presence of a laminal gesture (g071a019, cf. note 2)).

In one case a progressive assimilation of the place of articulation is observed. The initial /d/ is produced as a velar plosive after the velar plosive in *Dienstag, den* (g094a012). In German, progressive assimilation most often occurs in canonical plosive-schwa-nasal sequences where the schwa is not realised and the following nasal adopts the place of articulation of the plosive. Whereas these assimilations occur within words, the present case is unusual in that the assimilation crosses word- and phrase-boundaries.

The initial element of the definite article may be nasalised following a nasal segment. Such nasalisation may cooccur with an approximant articulation, or with complete oral closure resulting in [n]. An alveolar nasal (*d-n*) is marked in about 6% of cases. Furthermore, the replacement of the alveolar plosive by a nasal depends on the word class of the preceding item: it seems only to follow prepositions. The high rate of nasalisation for *der* (10%) results from the frequent occurrence of *in der* in the corpus. However, certain speakers (e.g. KAK) never produce a nasal stop instead of the oral one in *in der* sequences. Rather than being an automatic process, nasalisation is controlled by the speaker.

Fricativisation, where an alveolar fricative appears instead of a plosive, is conditioned by a preceding alveolar fricative, and is comparatively frequent for *das*. The alveolar fricative in which this item ends can regressively influence the articulation of the onset, which creates ideal conditions for fricativisation when additional alveolar friction is present before the onset.

The items discussed so far can have grammatical functions other than that of definite article, i.e. demonstrative and relative pronoun. In the absence of a reliable parser to assign grammatical function to a large number of tokens, the assigning was carried out manually for the subset of 6 speakers. For *der*, *die*, *dem*, and *den*, the vast majority of the tokens are definite articles (96, 92, 100, and 99%, respectively). Extrapolating these results to the larger corpora, it seems justified to talk about definite articles when referring to these items. The picture for *das*, however, is different: in 83% of cases it is used as a demonstrative pronoun ('that'), which accounts for the large number of tokens in spontaneous speech in comparison to the small number in the read corpus where the item is mainly used as an article.

When comparing the findings for spontaneous speech with those for read speech, the most important difference is that in the latter, there are more standard productions involving a plosive and its release, and fewer nasals instead of alveolar plosives.

3.2.2. Productions of final nasals. As for the forms of the indefinite article ending in a nasal (cf. 3.1.2.), the final nasals in *den* and *dem* can undergo assimilation of place and deletion. For *den* in spontaneous speech, there are 29 cases of assimilation of place. In 16 tokens there is a velar nasal before words beginning with velar, and in 13 productions the nasal is bilabial in bilabial contexts. Before alveolar nasals, the final nasal in *den* is either replaced by nasalisation of the preceding vowel (7 tokens) or completely dropped (9 tokens). Summing the above productions, connected speech phenomena affect the final nasal in 45 out of 646 tokens (7%). For read speech, 13 out of 140 tokens (9%) display symbolical modifications of the final nasals, with ten examples of articulation of place, and 3 tokens involving deletion of the nasal or nasalisation of the preceding vowel as a replacement for the contoid nasal articulation.

For *dem* in spontaneous speech, there are 11 out of 140 tokens (8%) involving modification of the final nasal, all of which are assimilations of place. In one token, there is a velar instead of a bilabial nasal, and in the remaining ten cases an alveolar nasal has been labelled. Some of the latter cases, however, are ambiguous in that it is not clear whether there is an assimilation of place, or whether the word used is actually *den* (cf. e.g. *Mittwoch, dem neunundzwanzigsten*, which might as well be *Mittwoch, den neunundzwanzigsten* (g072a016); both meaning 'Wednesday, the twenty-ninth'). No case of assimilation or deletion of the final nasal in *dem* was found in the 52 read tokens.

4. CONCLUSIONS

The findings suggest that traditional accounts of the phonetics of function words are at best incomplete, and in the case of glottalisation in connection with initial vowels, are inadequate. In both spontaneous and read speech, vowel onset is marked by creak rather than glottal closure (in accordance with [3]). Indeed, a glottal stop without creaky voice is actually the least common realisation. As for the definite article, the traditional description of an alveolar plosive as the standard variant is indeed accurate, but has to be complemented with the variants involving approximation, nasalisation, and fricativisation.

For the indefinite articles in spontaneous speech, there is a

connection between glottal activity at the onset of the vowel and phenomena induced by following articulations in the final nasals. In only 11% (2 out of 18) of these cases of assimilation and reduction does the token contain a glottal stop. The glottal stop is thus less frequent than for all tokens of the indefinite article in spontaneous speech (23%, cf. Table 1). The picture for read speech is rather different: 8 out of 31 cases (26 percent) of assimilation and reduction of the final nasal are produced with a glottal stop. This is even slightly more than the 21 percent of all tokens involving a glottal stop (note there is no pause before the tokens, which could explain the frequency of the glottal stop). In spontaneous speech there is thus a tendency for assimilation and reduction of the final nasals to be connected with the 'loss' of initial glottal stops, while this tendency does not exist in read speech. In other words, spontaneous speech exhibits greater homogeneity than read speech in the degree of different connected speech phenomena within the same word. So far, this finding only holds for the indefinite article; the /d/ in the definite article does not display more reduced correlates in connection with reduced forms of the final nasals.

We have shown that there is more laryngeal activity in spontaneous productions of the indefinite articles than in read ones because of the disfluencies of spontaneous speech, which impede the stronger form of reduction (absence of glottal reflex) found more frequently in read speech. On the other hand, approximants and nasals for /d/ in the definite articles are more frequent in spontaneous than in read speech. There is thus no simple relation between the speaking situation and the frequency of connected speech phenomena.

ACKNOWLEDGMENTS

Part of the work reported here was carried out with financial support from the German Research Council (DFG) within the project "Sound patterns of German spontaneous speech" (contract Ko 331/22-1). I would like to thank Klaus Kohler, the director of the project, Jonathan Rodgers, and Adrian Simpson for valuable comments and suggestions.

NOTES

REFERENCES

- [1] Mangold, Max (ed.) 1990. *Duden Aussprachewörterbuch* (3rd edition). Duden vol. 6. Mannheim, Wien, Zürich: Dudenverlag.
- [2] Krech, E.M. et al. (eds.) 1982. *Großes Wörterbuch der deutschen Aussprache*. Leipzig: VEB Bibliographisches Institut.
- [3] Kohler, K.J. 1994. Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica*, 51, 38-51.
- [4] IPDS 1995. *The Kiel Corpus of Spontaneous Speech Volume I*. Kiel-CD #2.
- [5] IPDS 1996. *The Kiel Corpus of Spontaneous Speech Volume II*. Kiel-CD #3.
- [6] IPDS 1997. *The Kiel Corpus of Spontaneous Speech Volume III*. Kiel-CD #4.
- [7] Kohler, K.J., Pätzold, M. and Simpson, A. 1995. From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech. *AIPUK* 29.
- [8] Kohler, K.J., Pätzold, M. and Simpson, A. 1997. From the acoustic data collection to a labelled speech data bank of spoken Standard German. *AIPUK*, 32, 1-29.
- [9] IPDS 1994. *The Kiel Corpus of Read Speech Volume I*. Kiel-CD #1.
- [10] Pätzold, M. 1997. *Kieldat* — Data bank utilities for the *Kiel Corpus*. *AIPUK*, 32, 117-126.

- 1) [aɪnəm] → [aɪnm] → [aɪmm] → [aɪm]
- 2) [nɔʏβəʔ]