Phonetic Exponents of Disfluency in German Spontaneous Speech*

Klaus J. Kohler, Benno Peters & Thomas Wesener

* This is a revised version of a paper first presented at Diss 01 - Disfluency in Spontaneous Speech and published under the title Interruption Glottalization in German Spontaneous Speech, in Proc. of Diss01 (2001), pp. 45-48.

Graphic signal representation and speech output of representative examples can be found at the following URL: www.ipds.uni-kiel.de/kjk/publikationen/audiobsp.en.html

We would like to thank Michel Scheffers for writing programmes that helped us in the automatic processing of the corpus data.
1 Introduction

Disfluencies have been studied extensively in the past decade, to a large extent in the context of speech technology application, where the aim is to filter out syntactic irregularities for more efficient automatic speech recognition (Bear, Dowding and Shriberg 1992; Nakatani and Hirschberg 1994). In this respect, the detection of repairs plays an important role, especially when it can be related to phonetic cues, such as glottalization.

The data base for this paper originated in a similar environment (Karger and Wahlster 1995). In the manual for orthographic transliteration of German dialogues (Kohler, Pätzold and Simpson 1995), two categories of syntactic irregularities were distinguished and marked symbolically: false starts and truncations. False starts refer to sentences that are broken off and continued after a simple repetition or repair (of a part) of what has already been said. For future filtering, the left and right edges of the reparandum are symbolized. In the case of a truncation, a sentence is broken off and not continued, but a new sentence may be started by the same speaker. In this case only the cut-off point is symbolized. A corpus of German dialogues annotated in this way is the point of departure for the basic research into disfluencies and their phonetic exponents, especially glottalization, presented in this paper. The aim is, however, that the results from this investigation will in turn be channelled into speech technology applications.

A preliminary study of glottalization phenomena as cues to false starts and truncations in German was carried out in the first phase of the DFG project ‘Sound patterns of German spontaneous speech’ (Rodgers 1999). The term ‘truncation glottalization’ used there is replaced here by ‘ interruption glottalization’ to cover glottalization in both truncations and false starts unambiguously, and to follow Nakatani and Hirschberg’s terminology (Nakatani and Hirschberg 1994).

In the following we are going to discuss the analysis of glottalization beside other phonetic exponents of false starts and truncations. ‘Glottalization’ refers to low frequency glottal pulsing (variable in frequency, amplitude and wave form; cf. Figure 1), in alternation with, or in addition to, glottal stop (cf. Figure 2). A further glottalization phenomenon is ‘tight voice’, which is characterized by a jump-up in F0 and by low amplitude, as well as by the auditory impression of tightness. Glottalization, glottal stop and tight voice are collectively referred to as glottalization phenomena. Other phonetic cues to these syntactic irregularities include interrup-
Figure 1: Glottalization in */ku=/* die Woche<Z> ‘* – the week’.

Figure 2: Glottal stop in ja, aber da w=/* ja<Z> ‘yes, but then * – yes’.
tions, within lexical material, by pauses, breathing, articulatory and non-articulatory noises, hesitation particles, and hesitational lengthening.

It is known from previous studies on English that glottalization frequently marks the end of reparandum intervals if they end in vowels (Bear, Dowding and Shriberg 1992; Nakatani and Hirschberg 1994). Our investigation picks up this thread and inserts it into the wider context of phonetic interruption cues generally.

Irregular glottal pulsing also functions as a signal of phrase finality without disfluency. This feature is associated with terminal pitch patterns that descend to the bottom of the pitch range, and plays no role in other contours. Therefore further data are presented on phonation in phrase-final terminal prosodies (modal voice versus irregular glottal pulsing) as one phonetic cue of phrase finality. For the sake of terminological stringency, this deviation from modal voice is called ‘laryngealization’ as opposed to interruption ‘glottalization’. Finally, the glottal cues of interruption and of phrase-finality are compared as to their frequency of occurrence and temporal extension.

2 Method

2.1 Database

The investigation reported in this paper is based on the Kiel Corpus of Spontaneous Speech (IPDS 1995, 1996). These data were collected in an appointment-scheduling scenario between two speakers who opened their own (and simultaneously closed their dialogue partner’s) recording channel by pressing a button (Kohler, Pätzold and Simpson 1995). These recordings are, therefore, not appropriate for analyzing disfluencies triggered by speaker interaction in overlapping dialogue. For the signal files manual transliterations were produced, including words in standardized orthography, the marking of pauses, articulatory and non-articulatory noises (e.g. breathing, paper rustling), hesitation phenomena, and syntactic irregularities.

Syntactic irregularities comprise deviations from syntactic structure, morphology, and lexicon. Those phenomena that are commonly known as slips of the tongue are included if they lead to syntactic irregularities. In the marking of syntactic irregularities, four types are distinguished: word-internal or word-external false starts (=/+ or /+) and word-internal or word-
external truncations (=/– or /–).

The transliteration files are automatically converted to transcription files containing canonical segmental word transcriptions as well as labels for the types of disfluencies mentioned above. These transcription files are the basis for the segmental labelling of the speech waves resulting in label files. The alphabet used is modified SAMPA.

Prosodic labelling is added to the label files and is done within the framework of the *Kiel Intonation Model* (KIM) (Kohler 1991). The point that is relevant for this presentation is the marking of low terminal pitch contours as &2., followed by a prosodic phrase boundary label.

Only the data files with complete segmental and prosodic labelling are used for this investigation. This corresponds to the signal files of all complete sessions in volumes 1 and 2 of the *Kiel Corpus of Spontaneous Speech*, with 22 speakers (13 male, 9 female) in 11 dialogue sessions; the total recording time amounts to approximately 2.5 hours (25000 words). The label files of this selected data base are entered into a structured data bank using *Kieldat* utilities (Pätzold 1997).

### 2.2 Data search and data processing

The data search is carried out with reference to the marking of

- syntactic irregularities (=/+ , /+, =/–, /–), data set A
- low terminal falls at prosodic phrase boundaries (&2.), data set B

Syntactic irregularity, represented by data set A, is the initial criterion in the search for disfluencies. It may not be coupled with phonetic exponents signalling an interruption. If there is a phonetic manifestation, it may be a glottal stop, glottalization, tight voice, pausing, breathing, a hesitation particle, or hesitational lengthening, which also includes holding a stop closure, and possibly others.

The selected data base is searched for all occurrences of syntactic irregularities (data set A), and low terminal falls (data set B) in a frame from the prosodic phrase boundary preceding to the one following any one of the respective labels. The signal portions corresponding to each of these data sets are automatically spliced together. In parallel to the signal file for each data set, two text files are generated, providing (1) the orthographic words and (2) the segmental and prosodic labels and their time points.
In the case of data set A, the next step is the automatic extraction, from
the data bank, of the type of irregularity marker, the segmental contexts im-
mmediately preceding or following, and the speaker identification. For data
set B, the automatically extracted information refers to the labels preceding
and following the terminal contour marker as well as to the speaker iden-
tification. In both data sets, the preceding context is classified as sonorant
(vowel, nasal, lateral) or non-sonorant, the following context as phonolog-
ical segment or canonical glottal stop or pause/breathing or other (articula-
tory or non-articulatory) noises. A new label containing the automatically
extracted information is introduced in the label files (2) at the time point of
each syntactic irregularity or prosodic contour marker, respectively.

Canonical glottal stop refers to the automatic transcription of a glott-
tal stop symbol before all word-initial vowels in German. This glottal
stop may be realized as such or as glottalization, or not at all. If there
is such a canonical glottal stop following a point of syntactic irregular-
ity or a phrase-final terminal prosody, the actual occurrence of a glottal
stop or glottalization may be ambivalent in its reference either to pho-
etic interruption/phrase-final laryngealization, or to the following canoni-
cal glottal stop. These ambivalent cases are excluded from further analysis.

The files for each data set are then analyzed by accessing speech wave,
spectrogram, fundamental frequency, orthographic words, and labels in
parallel windows with the xassp programme (IPDS 1997). As regards data
set A, the automatically generated classifiers are manually supplemented
by adding information on the phonation type at the end of the reparan-
dum, with the four-fold specification of glottal stop or glottalization or
tight phonation or their absence; an additional classifier is reserved for un-
certain cases. The respective label is added to the automatically inserted
label string.

As regards data set B, the information on phonation type, which is
added manually to the automatically generated classifiers, provides the
five-fold specification of glottal stop or laryngealization or modal voice
or plosive-related glottalization or uncertain. Plosive-related glottalization
refers to the realization of plosives as glottal stop or glottalization in, e.g.,
bilateral nasal environment (könnten [kʰøntən] ‘could’ (Kohler 2001)). If
a word with such a phonotactic structure occurs in a terminal fall at the
end of a prosodic phrase, the incidence of irregular glottal pulsing cannot
be uniquely associated with phrase-final laryngealization; therefore these
cases are excluded from further analysis.
In order to determine the temporal extension of glottalization and laryngealization, the stretches of irregular glottal pulsing are manually labelled with begin and end markers for both data set A and B.

3 Results

3.1 Phonetic cues at false starts and truncations

3.1.1 Description

There are 338 instances of marked syntactic irregularities. Of these, 41 are followed by a canonical glottal stop, 17 are labelled as uncertain phonation; these cases are excluded from further analysis. On the basis of the automatic classifications of the phonetic environments of syntactic irregularity markers and the manual classifications of the phonation types at these places, the following categories of phonetic interruption within lexical material were defined:

- interruption by glottal stop (class I) or glottalization (class II) or tight voice (class III)
- interruption by pause or breathing or (non)-articulatory noises (class IV)
- interruption by hesitation particles (class V)
- interruption by hesitational lengthening (class VI)
- no interruption (class VII)

In the case of multiple cues for phonetic interruption, the following precedences determine classification: glottalization phenomena (classes I–III) over pause/breathing/(non)-articulatory noises (class IV) and hesitational lengthening (class VI); pause/breathing/(non)-articulatory noises (class IV) over hesitational lengthening (class VI); hesitation particles (class V) over glottalization phenomena (classes I–III) and hesitation lengthening (class VI). If there is a sequence of class IV and class V features, the one that comes first determines classification.

Looking at the covariance between the four defined classes of syntactic irregularities — word-internal/word-external false starts, word-internal/word-external truncations — and the seven classes of phonetic
Table 1: Covariance between classes of syntactic irregularities (=/+ word-internal false start, /+ word-external false start, =/– word-internal truncation, /– word-external truncation) and classes of phonetic interruption (I–VII).

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>=/+</td>
<td>14</td>
<td>17</td>
<td>4</td>
<td>15</td>
<td>9</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>/+</td>
<td>7</td>
<td>12</td>
<td>4</td>
<td>43</td>
<td>5</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>=/–</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/–</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>37</td>
<td>7</td>
<td>3</td>
<td>12</td>
</tr>
</tbody>
</table>

interruption across the total speaker population we arrive at the data distribution presented in Table 1. Glottalization phenomena, summed over the classes I–III, mark 27% of all cases of syntactic irregularities; there is no difference in their distribution across the preceding sonorant or non-sonorant contexts. This relative frequency is practically identical with the 26% in class VII, which has no phonetic interruption cue (cf. Figure 3). The highest proportion (35%) is associated with phonetic interruption by pause or breathing or (non)-articulatory noises. If classes IV and V are conflated, we get 43% non-verbal insertions.

Internal truncations have low frequency, whereas the other classes of syntactic irregularities occur more frequently.

Figure 4 presents the data of Table 1 with classes I–III grouped as glottal cues, and classes IV–VI as other cues. Word-internal and word-external irregularities show opposite distribution patterns. Word-internal false starts and truncations show a high incidence of glottal phenomena and of the absence of a phonetic cue, at the expense of other cues; in word-external false starts and truncations, these distributions are reversed.

The mean duration of irregular glottal pulsing in cases of glottalization (class II) is 76 ms.

An examination of the behaviour of the individual speakers shows diverging trends between them. On the one hand there are speakers who have very few or no glottalizations in any of the four types of syntactic irregularities, and especially use breathing instead. On the other hand, the distribution across the four types of syntactic irregularities differs a great deal from speaker to speaker. This means that the group data may be biassed by individual speakers, particularly since the total frequency of syntactic irregularities per speaker is not very high and differs from speaker to speaker. For this reason the data presentation is broken down into the
frequency distributions of individual speakers with the four categories of syntactic irregularities conflated into one. This is done for those speakers who produced more than 15 cases (maximum 31) in Table 2.

Three speakers (CHD, ANL, HAH) have no glottalization phenomena, and at the same time show the highest frequency in the category of non-verbal insertions (classes IV and V). Two speakers (TIS, FRS) show the opposite trend. Figure 5 visualizes these two strategies for speakers TIS and ANL. The remaining three speakers fall in between these two groups.

3.1.2 Interpretation

The following tentative interpretation is offered for the low frequency of word-internal truncations. In the absence of a proper dialogue situation in the recording scenario there is no overlap between speakers, so a speaker is not compelled to stop at a time when the other speaker starts speaking, which may be at any point in verbal material. Interruptions internally in a speaker’s turn, on the other hand, may be predominantly of the false start type, and if they are of the truncation type they may not occur before the end of a word is reached.
The facts that the highest frequency of glottal phenomena occurs in internal false starts and that word-internal and word-external false starts show opposite distribution patterns may be seen as indicating a reinforcement of the fragment nature of the verbal material, whereas non-verbal insertions seem to be used to mark interruptions at word boundaries. This evaluation of the German data can be connected with the report by Nakatani and Hirschberg (Nakatani and Hirschberg 1994) that the majority of the glottalizations they found in English occur in word-fragments.

Table 2: Classes of phonetic interruption (I–VII) for different speakers. The four categories of syntactic irregularity are merged into one.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIS</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>FRS</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>JAK</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>SAR</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>OLV</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>CHD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>ANL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>4</td>
<td>1</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td>HAH</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>21</td>
</tr>
</tbody>
</table>

15 20 4 75 12 10 48 184
This interpretation, however, has to be taken with caution because when we look at the behaviour of different speakers we find diverging trends. There are speakers who do not seem to use glottalization phenomena as phonetic interruption cues but have a preponderance of non-verbal insertions, and there are others for whom the reverse applies (Fig. 5). So we have to take speaker-specific preferences into consideration.

The analysis results of interruption glottalization in the German corpus is comparable to English data discussed by Nakatani and Hirschberg, who found 30% of all reperanda offsets to be glottalized (38% in word-internal position). Bear et al. (1992), however, found a much higher proportion of fragments with glottal cues (24 out of 25 analyzed cases that end in a vowel).

In our data, the frequency of the absence of interruption cues is about as high as the frequency of glottal cues. This means that although speakers may mark a syntactic irregularity by an abrupt phonetic cut-off in order to signal to the hearer that they are, e.g., going to correct themselves, they may also do the precise opposite and gloss over their false starts and truncations. So we should adopt a more differential view of the link between interruption glottalization and syntactic irregularities in that the use of different interruption cues or their absence may be related to changing intentions and/or situational constraints in one speaker, or it may characterize different speakers’ behaviours.
Table 3: Frequency distribution of laryngealization and modal voice across three types of phrase finality: TF turn-final, TINV turn-internal before non-verbal material, TIV turn-internal before verbal material.

<table>
<thead>
<tr>
<th></th>
<th>TF</th>
<th>TINV</th>
<th>TIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>laryngealized</td>
<td>119</td>
<td>76</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>57%</td>
<td>23%</td>
<td>29%</td>
</tr>
<tr>
<td>modal voice</td>
<td>89</td>
<td>256</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>43%</td>
<td>77%</td>
<td>71%</td>
</tr>
<tr>
<td></td>
<td>208</td>
<td>332</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>68%</td>
<td>68%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1111</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Phrase-final laryngealization

There are 1633 instances of low terminal pitch contours at prosodic phrase boundaries. Of these, 127 can be connected with plosive-related glottalization, 205 with the phonetic realization of a following canonical glottal stop, and 180 are uncertain. This leaves 1121 cases for further analysis. Among these, 752 have modal voice, 359 laryngealization, and 10 end in a glottal stop. The mean duration of irregular glottal pulsing in cases of laryngealization is 164 ms.

The occurrence of a glottal stop in this prosodic position is negligible. Table 3 gives the distribution of the phonetic cues of laryngealization and modal voice, respectively, across the three types of phrase-finality: turn-final, turn-internal before pauses/breathing/(non)-articulatory noises, and turn-internal before verbal material. It shows a much higher incidence of laryngealization turn-final than turn-internal, and almost identical distributions of laryngealization and modal voice in the two internal types.

Like interruption glottalization, phrase-final laryngealization also shows speaker-specific behaviour. There are speakers who have laryngealization quite regularly and others that have very few cases. Among the 22 speakers, 4 have relative frequencies below 10% (one 0%), 7 between 10 and 30%, 9 between 40 and 60%, and 2 have 70 and 72%, respectively. The group data, in conjunction with these individual distributions, suggest that speakers use laryngealization predominantly at the end of turns, and have diverging preferences for the use of this additional phonetic marker of terminal phrase finality.
4 Discussion

Interruption glottalization and phrase-final laryngealization differ in several respects:

a) Interruption glottalization includes the glottal stop quite frequently, laryngealization does not.

b) Interruption glottalization is associated locally with the point of interruption and sounds tense, whereas final laryngealization is realized over longer stretches (76 ms vs 164 ms), and sounds lax.

c) There are also differences of spectral characteristics between the two phenomena.

d) Laryngealization is always associated with low falling F0, glottalization occurs at the level in the F0 contour that has been reached at the utterance break, which is often high.

The impressionistic observations in c)–d) need further systematic quantification as regards differences between glottalization and laryngalization in spectrum and intensity (c), and in F0 context (d).

An important finding of this investigation is that both syntactic irregularities and phrase finality are signalled by multiple acoustic cues which are used in different combinations by individual speakers. Glottalization phenomena are optional markers in addition to, or instead of, other phonetic interruption features, and laryngealization is optional in addition to low terminal F0 and phrase-final lengthening. Both glottalization and laryngealization provide a strengthening of the respective signals for utterance breaks and phrase finality, and in the latter case, the turn-final position is given extra prominence. The cases for which no phonetic interruption has been recorded at syntactic irregularities require more detailed signal analysis to see whether a special pitch feature, e.g. a high F0 onset after the point of syntactic irregularity (as found in English by Bear et al. (1992)), still cues a break, albeit more weakly. It must also be pointed out that classification into ‘laryngealized’ and ‘modal voice’ is very coarse, it needs further refinement into breathy beside modal voice and breathy laryngealization beside laryngealization (and possibly other phonation types).

One of the semantic functions of truncation glottalization is for speakers to indicate that they change plan and want to hold their turn before
repairing or starting a new utterance. However, glottalization also occurs when a speaker is interrupted by another speaker, or attempts to take a turn without succeeding (Local and Kelly 1986). A subsequent investigation with further material from overlapping dialogues (compared with the non-overlapping ones of the Kiel Corpus) will be carried out to analyze these glottalizations in turn-holding and turn-taking strategies (Peters 2005).
References


Pätzold, M. (1997). KielDat - data bank utilities for the Kiel Corpus. In A. Simpson, K. J. Kohler, and T. Rettstadt (Eds.), The Kiel Corpus of
Read/Spontaneous Speech: Asoustic data base, processing tools and analysis results, AIPUK 32, pp. 117–126. Kiel: IPDS.
