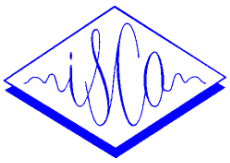

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Interruption glottalization in German spontaneous speech

Klaus J. Kohler, Benno Peters, Thomas Wesener

Institute of Phonetics and Digital Speech Processing
University of Kiel, Germany
kk@ipds.uni-kiel.de

Abstract

This paper analyzes the occurrence of phonetic interruption cues at points of syntactic irregularities (false starts and truncations) in a large annotated corpus of German dialogues and compares interruption glottalization with laryngealization in terminal low phrase-final prosodies. Glottalization (including glottal stop) predominantly marks word fragments, whereas non-verbal insertions, e.g. breathing, tend to be word-external interruption cues. Laryngealization (excluding glottal stop) predominantly signals terminal phrase boundaries in turn-final positions. Individual speakers differ a great deal as to the distribution of these phenomena.

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 [1] [2]. 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 [3]. In the manual orthographic transliteration of German dialogues, 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 presented in [4]. 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 [2].

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), in alternation with, or in addition to, glottal stop. 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 interruptions, within lexical material, by pauses, breathing, articulatory and non-articulatory noises, hesitation particles, and hesitational lengthening[1].

It is known from previous studies on English that glottalization frequently marks the end of reparandum intervals if they end in vowels [1] [2]. 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 phonetic cues of interruption and of phrase-finality are compared.

2. Method

2.1. Data base

The investigation reported in this paper is based on the *Kiel Corpus of Spontaneous Speech* [5]. 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 [6]. 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 transcrip-

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

tions 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) [7]. 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 [8].

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 with *awk* scripts 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 immediately 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 identification. In both data sets, the preceding context is classified as sonorant (vowel, nasal, lateral) or non-sonorant, the following context as phonological segment or canonical glottal stop or pause/breathing or other (articulatory 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 glottal 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 irregularity or a phrase-final terminal prosody, the *actual* occurrence of a glottal stop or glottalization may be ambivalent in its reference either to phonetic interruption/phrase-final laryngealization, or to the following canonical 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 [9]. As regards data set A, the automatically generated classifiers are manually supplemented by adding information on the phonation type at the end of the reparandum, with the four-fold specification of glottal stop or glottalization or tight phonation or their absence; an additional classifier is reserved for uncertain 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* [kʰœn̩n̩n] 'could' [10]). 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.

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, on the one hand, and the manual classifications of the phonation types at these places, on the other, a category of presence/absence of phonetic interruption within lexical material was defined in the following way:

- 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 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%

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).

	I	II	III	IV	V	VI	VII	
=/+	14	17	4	15	9	3	37	99
/+	7	12	4	43	5	6	22	99
=/–	1	1	1	4	0	1	1	9
/–	8	3	3	37	7	3	12	73
	30	33	12	99	21	13	72	280

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

	I	II	III	IV	V	VI	VII
TIS	4	7	3	3	1	2	7
FRS	4	3	1	2	2	1	4
JAK	2	6	0	14	0	1	4
SAR	5	0	0	11	0	0	6
OLV	0	4	0	3	3	1	9
CHD	0	0	0	13	0	0	6
ANL	0	0	0	17	4	1	9
HAH	0	0	0	12	2	4	3
	15	20	4	75	12	10	48

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 of any kind. 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.

Word-internal and word-external false starts show opposite distribution patterns. The former have a high incidence of the absence of a phonetic cue, as well as of glottalization phenomena, at the expense of non-verbal insertions; in the latter category, these distributions are reversed. Word-external false starts and truncations show similar patterns. Two lines in table 1 present extreme distributions: 1) Internal truncations have very low frequency. 2) Internal false starts have the highest frequency of glottal phenomena.

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 26) 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. 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 [2] that the majority of the glottalizations they found in English occur in word-fragments.

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. So we have to take speaker-specific preferences into consideration.

The analysis results of interruption glottalization in the German corpus differ in two important respects from the English data discussed in the literature. The relative frequency of glottalization overall and in word-fragments in particular is lower [1] [2], and the frequency of the absence of interruption cues is about as high as the frequency of glottalization. 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, on the one hand, or characterize different speakers' behaviours, on the other.

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 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.

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.

	TF	TINV	TIV	
laryngealized	119	76	164	359
	57%	23%	29%	32%
modal voice	89	256	407	752
	43%	77%	71%	68%
	208	332	571	1111

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, 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 b)–d) need systematic quantification as regards the extension of glottalization and laryngealization over time and number of sound elements (b), their differences in spectrum and intensity (c), and 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. [1]), 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 [11]. A subsequent investigation with further material from overlapping dialogues (compared with the non-overlapping ones of the Kiel Corpus) will be necessary to analyze these glottalizations in turn-holding and turn-taking strategies.

5. Acknowledgements

Part of the work reported here was funded by German Research Council Grant Ko 331/22-2 ("Sound patterns of German spontaneous speech"). We would also like to thank Michel Scheffers for writing programmes that helped us in the automatic processing of the corpus data.

6. References

[1] Bear, J. and Dowding, J. and Shriberg, E., "Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog", Proceedings of the 30th Annual Meeting (Association for Computational Linguistics), Newark, DE, 1992, 56–63.

[2] Nakatani, C. and Hirschberg, J., "A corpus-based study of repair cues in spontaneous speech", J. Acoust. Soc. Amer., Vol. 95, 1994, 1603–1616.

[3] Karger, R. and Wahlster, W., Verbmobil Handbuch – Version 3, Verbmobil Technisches Dokument Nr. 35, DFKI, Saarbrücken, 1995.

[4] Rodgers, J. E. J., "Three influences on glottalization in read and spontaneous German speech", AIPUK, Vol. 34, 1999, 177–284.

[5] IPDS, The Kiel Corpus of Spontaneous Speech Vol. 1–3, Institute of Phonetics and Digital Speech Processing, Kiel, 1995–1997.

[6] Kohler, K. J., Pätzold, M., and Simpson, A. P., From scenario to segment: the controlled elicitation, transcription, segmentation and labelling of spontaneous speech, AIPUK, Vol. 29, 1995.

[7] Kohler, K. J., "A model of German intonation", AIPUK, Vol. 25, 1992, 295–360.

[8] Pätzold, M., "*KielDat* – Data bank utilities for the *Kiel Corpus*", AIPUK, Vol. 32, 1997, 117–126.

[9] IPDS, *xassp* User's Manual (Advanced Speech Signal Processor under the X Window System), AIPUK, Vol. 32, 1997, 31–115.

[10] Kohler, K. J., "Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German?", in Grønnum, N. and Rischel, J. (eds) To Honour Eli Fischer-Jørgensen, Travaux du Cercle Linguistique de Copenhague 31, C. A. Reitzel, Copenhagen, 2001, 174–211.

[11] Local, J. K. and Kelly, J., "Projection and 'silences': notes on phonetic detail and conversational structure", Human Studies, Vol. 9, 1986, 185–204.