Glottal filled pauses in German

Malte Belz
Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany

Abstract

For German, filled pauses are traditionally described with a vocalic form *äh* and a vocalic-nasal form *ähm*. A corpus-based approach and a closer phonetic inspection is used here to argue for an additional form, namely glottal filled pauses. In the data analysed for this study, the glottal form is produced by all seven speakers and amounts to 21% of all filled pauses. Contexts and durations of occurrences are discussed and compared to earlier studies on traditional filled pauses. It is suggested that the glottal variant should be considered in future studies on filled pauses and disfluencies.

Acoustic forms of filled pauses

Filled pauses (FPs) are defined in many ways. In this paper, I will use as a working hypothesis the notion of non-lexical entities, without considering extra-linguistic events (laughing, coughing, etc.). FPs are used as hesitation devices, but also serve other functions (Lickley 2015: 463). They are often exemplified with graphemic or phonetic realizations of the most frequent forms in a respective language. Lounsbury (1954) transliterates FPs in English as *hem* and *haw*, while Maclay and Osgood (1959) give more phonetic detail by listing the transcripts [ɛ æ r ə m]. It is largely agreed on that FPs often exhibit "both a prolonged vowel sound and a vowel (usually) followed by a nasal" (Lickley 2015: 458). Other forms that may be subsumed under FPs are clicks (Trouvain, Fauth & Möbius, 2015) and breath pauses (Trouvain et al. 2016). For German, the most cited forms are probably *äh* and *ähm*, with possible phonetic transcriptions varying between [ɛː] or [əː] and [ɛːm] or [əːm], although other forms are mentioned as well (cf. Schönle & Conrad 1985 for *ah* and *mh*).

In this paper, I will explore whether an additional form – a glottal filled pause – can be assumed for German spontaneous speech. This research question is part of a PhD project, in which I am currently investigating the link between form and function of FPs. In the process of annotation (cf. Section 2), I noticed sequences of glottal pulses and creak phonation without coarticulated vowels that seem to be used in a similar way to other FPs.

Example 1 gives a broad transcription of a speaker's utterance in a dialogue[i] of the GECO corpus (cf. Data and annotation). Durations are given within angle brackets. The speaker produces a glottalized sequence of approximately 16 glottal pulses before uttering another *yes* (cf. Figure 1a for a depiction of the signal). For a first description, a wildcard notation of a creaky sonorant is used.

(1) <[jaː] 580 ms> <*exhalation* 373 ms> <*inhalation* 718 ms> <[S̰ː][a] 492 ms> < [jaː] 630 ms> <[ç ˈvaɪs nɪ ˈalzoː] 706 ms>

a. S ≙ sonorant

Transliterations:
‚ja, ja ich weiß nicht also'
'yes, yes I don't know, well'

This glottalized sequence seems to be different from the rule-governed [ʔ]-epenthesis as predicted in German phonology:

(2) $\emptyset \rightarrow [ʔ] / \left\{ \begin{matrix} V \_ \ 'V \\ \# \_ \ V \end{matrix} \right\}$ (Hall 2011: 66)

Data and annotation

To investigate the forms of filled pauses, a multi-layer annotation scheme was added to the corpus GECO (GErman COnvergence) (Schweitzer & Lewandowski 2013). Six dialogues of the multi-modal condition are annotated as to now. In this condition, interlocutors are visible to each other, while speaking freely about any subject, separated by a transparent window and connected via headphones. Five of seven participants (A, C, K, M, D) participate in two dialogues. Each dialogue lasts 25 minutes. All speakers are female students, some with a noticeable Southern German (Swabian) accent.

Filled pauses are marked on a hesitation tier in Praat (Boersma 2001) with *fv* (vocalic or vocalic-nasal or nasal filler), *fg* (glottal filler) or *fc* (click filler), based on the perceptual categorization of the annotator. For the annotation of *fg*, additional cues were used such as irregular voicing periods (oscillogram) or clearly visible glottal stops (spectrogram). A further, yet preliminary approach is that no prominent vowel quality can be perceived.

The last speech segment to the immediate left and the first to the right of an FP are marked on the hesitation layer with *as* (antecedent segment), *ap* (antecedent silent pause), *ah* (antecedent breath pause), *ac* (antecedent click), *at* (antecedent turn), and postcedent *ps, pp, ph, pc* and *pt*, respectively. Further transcriptions of *a, f* and *p* categories and pause specifics (inhalation, exhalation) are annotated on a segmentation layer.

All files are converted to an EMU speech database (Winkelmann, Harrington & Jänsch, 2017) with help of the *emuR* package 0.2.1 (Winkelmann et al. 2016) in R (R Core Team 2016).

Results

Filled pause types

Table 1 shows the distribution of glottal, vocalic and click filled pauses per speaker.

Table 1. Filled pause type, word count and total frequency of FP per speaker.

	fc		fg		fv		Σ	Words	FP
	N	%	N	%	N	%	N	N	%
A	3	9.1	13	39.4	17	51.5	33	4214	.78
C	27	25.7	23	21.9	55	52.4	105	4766	2.2
K	9	7.6	31	26.1	79	66.4	119	8193	1.5
D	0	0	15	18.5	66	81.5	81	4987	1.6
F	0	0	1	16.7	5	83.3	6	1401	.43
M	0	0	8	11.0	65	89.0	73	5880	1.2
J	0	0	1	3.8	25	96.2	26	3092	.84
Σ	39	8.8	92	20.7	312	70.4	443	32533	1.4

Vocalic FPs are the most frequent form (70.4%), followed by glottal FPs (20.7%) and click FPs (8.8%). Some speakers do not utter any click FPs at all, whereas glottal FPs do occur at least once per speaker. The glottal FP with antecedent inhalation pause and postcedent *ja* (cf. Example 1) is shown in Figure 1a. An example with antecedent segmental and postcedent silent context is given in Figure 1b.

Immediate context

Figure 2 shows the distribution of antecedent, FP, and postcedent per FP type. The right tail of the distribution is cut off at five instances or less for plotting purposes, omitting 9.4% of the data. In the first bar of Figure 2 (*as_FP_ps*), 23.2% of all contexts with adjacent speech segments are glottal FPs. In the second and third bar, contexts with antecedent silent pauses and segmental postcedents (*ap_FP_ps*) exhibit more glottal FPs than those with antecedent breath pauses (*ah_FP_ps*).

This difference between silent and breath pauses is furthered when the contexts are reversed. In the fourth and fifth bar of Figure 2, contexts with segmental antecedents and postcedent silent pauses (*as_FP_pp*) exhibit glottal FPs, whereas those with postcedent breath pauses (*as_FP_ph*) show no glottal FPs at all. Most of the breath pauses considered here are inhalation breath pauses (50 of 52 in *ah_FP_*ps and 27 of 36 in *as_FP_ph*).

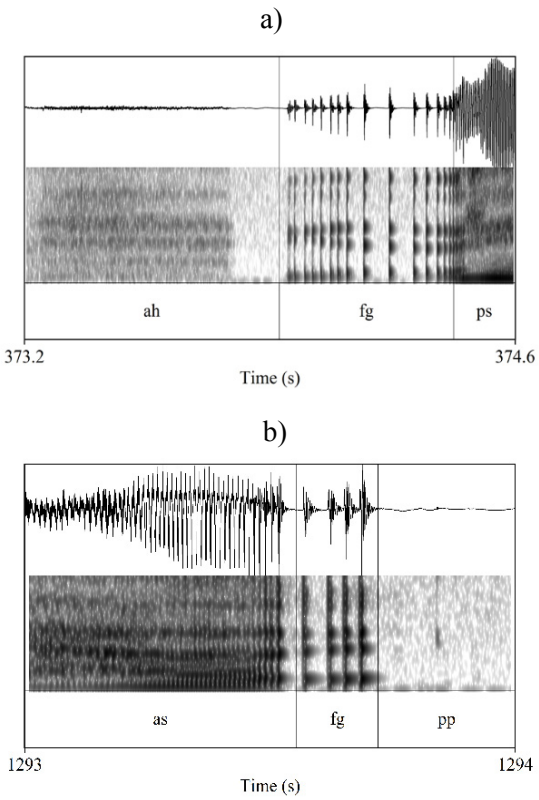


Figure 1. ***a)*** *Glottal filled pause (492 ms) labelled* fg*, preceded by breath intake (718 ms) and followed by [j] (173 ms) in* ja 'yes' *of subject C speaking with D.* ***b)*** *Glottal filled pause (88 ms) labelled* fg*, preceded by [a:] (296 ms) in* ja 'yes' *and followed by a silent pause (149 ms) of subject C speaking with D.*

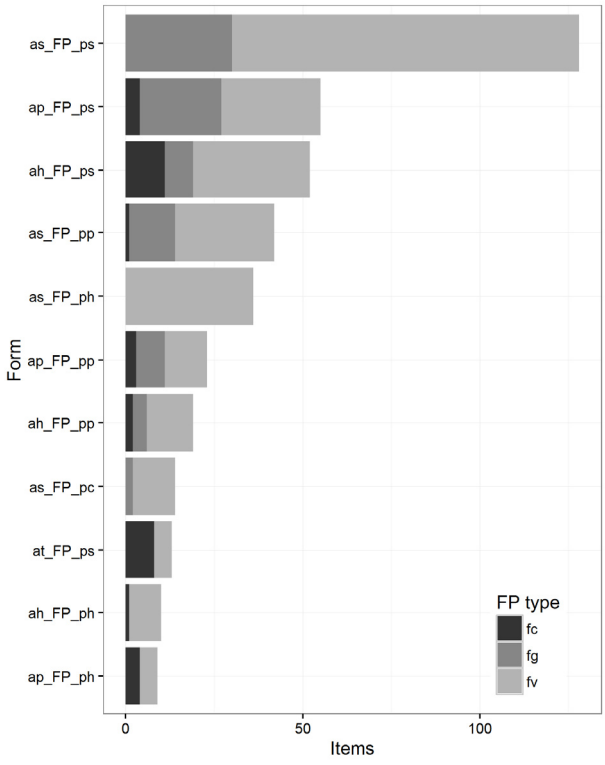


Figure 2. Distribution of filled pause forms per type. 'FP' *is a substitution for any type of* fc*,* fg*, or* fv*.*

Durational features

Vocalic and vocalic-nasal form show the longest durations, click FPs the shortest. This is plausible, given the articulatory features of clicks. The durations of glottal FPs are in between. Table 2 presents the mean lengths and standard deviations of FP types. Each paired comparison is significant (*fc* vs. *fg*: $t = 5.4$, $df = 106$, $p < .001$; *fc* vs. *fv*: $t = 14$, $df = 61$, $p < .001$; *fg* vs. *fv*: $t = 6.2$, $df = 140$, $p < .001$). The lengths of the vocalic and vocalic-nasal forms of *fv* (*äh* and *ähm*, without a closer inspection with respect to their vowel quality) also differ significantly ($t = -7.8$, $df = 233.3$, $p < .001$). However, the difference between glottal FPs (*fg*) and the *fv* variant *äh* is not significant ($t = -1.6$, $df = 163$, $p = .1$).

Table 2. Measures of central tendencies for vocalic fillers (fv), *glottal fillers* (fg), *and click fillers* (fc) *in milliseconds. Vocalic fillers are further split in vocalic-only and vocalic-nasal fillers.*

		x̄	SD
fc		109.7	86.1
fg		232.9	159.2
fv		352.9	151.2
fv	*äh*	267.1	125.9
fv	*ähm*	390.7	122.4

Glottal FP vs. laryngealization

Utterance-final or word-final glottalization due to a declined fundamental frequency (f_0) and the nearing minimum of air capacity in the lungs is sometimes called laryngealization (Kohler, Peters & Wesener, 2005: 189). How are instances of glottal FPs as in *as_fg_pp* different, then, from laryngealized speech? Figure 3 and Figure 1b give some qualitative evidence by comparing two within-speaker instances of the lemma *ja* 'yes'. In Figure 3, the vowel of *ja* is laryngealized towards the end. The glottal sequence in Figure 1b, however, is made out of four clearly perceivable single pulses and an interrupted voice bar.

Discussion and conclusion

Forms of FPs are language-specific (Clark & Fox Tree 2002: 92; Leeuw 2007; Wieling et al. 2016). However, two forms of FPs are ubiquitously mentioned in the literature for various vernaculars – a vocalic-only form (*uh* in English, *äh* in German), and a vocalic-nasal form (*uhm* and *ähm*, respectively). Perceivable breath pauses and clicks are sometimes also considered FPs. This paper argues for another type of FP – a glottal variant.

In spontaneous speech, the rule of [ʔ]-epenthesis for German (cf. Example 2) is not always met.

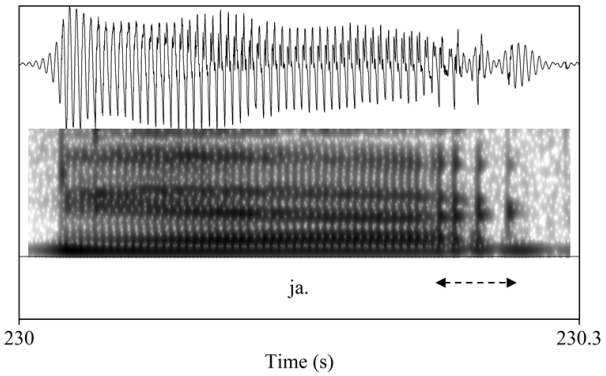


Figure 3. Example of laryngalized [jaː] ja 'yes' *(303 ms) uttered by subject C speaking with D. The laryngalized part lasts 75 ms and is indicated with the dashed arrows.*

Glottal marking of vowel-initial words in German also depends on speaking rate: the faster the rate, the less glottalization is observed (Pompino-Marschall & Żygis 2010). Nevertheless, the tendency for vowel-initial glottal stop insertion in German might add to the emergence of a glottal FP. The use of glottal stops as a functional marker is not a new phenomenon. After all, glottalization in German is also used to mark truncations (Kohler, Peters & Wesener, 2005).

Individual variation between speakers is a known challenge in FP research, and Table 1 gives ample evidence for that. For example, click FPs are only uttered by three speakers. However, although the speaker sample up to now is small and criticizable, the argumentation in favor of a new glottal FP category is strengthened by the fact that each of the speakers produces a glottal FP at least once.

Glottal FPs are found in many, but not all of the contexts where traditional vocalic and vocalic-nasal FPs occur. Even though their communicative function is yet unclear, speakers use them frequently. The non-occurrence of sequences consisting of a segmental antecedent, a glottal FP and a breath pause (*as_fg_ph*) might be related to the beginning inhalation process, in which the vocal folds are in abducted position, thus physiologically inhibiting the production of a continuing sequence of adduction gestures. A tentative implication of this physiological restriction is that speakers avoid glottal FPs in this context and produce vocalic or vocalic-nasal FPs before they run out of breath.

Strikingly, the durational distributions of glottal FPs and vocalic-only FPs overlap (cf. Table 2). One explanation is that glottal FPs consist of either one to three clearly perceivable glottal stops, or a larger sequence of creak phonation on top of an underspecified sonorant (in lack of a better description). This creaky sonorant can then be lengthened. The glottal FP, therefore, is either used as an allo-FP to *äh*, or speakers ascribe another function to it.

It seems that glottal FPs differ from non-FP word-final laryngealization and from coarticulated vowel-internal glottalization as in [ʔɛ̰ː]. Glottal FPs are, impressionistically, auditorily more prominent than word-final laryngealization and show higher glottal pulse energy. From a comparison of the sound pressure level of the laryngealized part of *ja* (Figure 3) with the glottal FP after (a different) *ja* (Figure 4) it seems that the glottal FP is uttered with a higher articulatory effort. However, a clear distinction between glottalized [ʔɛ̰ː] and a glottal FP with a conjectural form [ʢ] remains dubious. Further research will show whether they are used in a distinct or interchangeable way.

What are the merits from yet another (along with breath pauses and clicks) attested form of FPs?

First, we have to account for its mere observance.

Second, it has been shown that a more specific phonetic description apart from the assumption of standardized graphematic FP forms is a fruitful approach to reflect the actual variability in FP type, context and timing distributions more clearly.

Third, the analysis of contexts and glottal FPs might add to the debate of FPs being an intentional signal vs. FPs being an epiphenomenal, cognitive-burden induced entity (cf. Nicholson (2007) for an overview). At least the non-occurrence of glottal FPs in certain contexts might be explained by respiratory limitations.

References

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9). 341–345.

Clark, H. H. & J. E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1):73–111.

De Leeuw, E. 2007. Hesitation Markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19(2):85–114.

Hall, T. A. 2011. *Phonologie*: *Eine Einführung*. Berlin/New York: de Gruyter.

Kohler, K. J., B. Peters & T. Wesener. 2005. Phonetic Exponents of Disfluency in German Spontaneous Speech. In *Prosodic Structures in German Spontaneous Speech* (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel), 185–201.

Lickley, R. J. 2015. Fluency and Disfluency. In Melissa A. Redford (ed.), *The Handbook of Speech Production*, Hoboken, NJ: John Wiley & Sons, Inc., 445–469.

Lounsbury, F. G. 1954. Transitional Probability, Linguistic Structure, and Systems of Habit-family Hierarchies. In Charles E. Osgood & Thomas A. Sebeok (eds.), *Psycholinguistics*: *A survey of theory and research problems*. Baltimore: Waverly Press, 93–101.

Maclay, H. & C. E. Osgood. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 5:19–44.

Nicholson, H. B. M. 2007. *Disfluency in Dialogue: Attention, Structure and Function*. PhD Thesis, University of Edinburgh.

Pompino-Marschall, B. & M. Żygis. 2010. Glottal Marking of Vowel-Initial Words in German. *ZAS Papers in Linguistics*(52). 1–17.

R Core Team. 2016. *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.

Schönle, P.-W. & B. Conrad. 1985. Hesitation vowels: a motor speech respiration hypothesis. *Neuroscience Letters* 55:293–296.

Schweitzer, A. & N. Lewandowski. 2013. Convergence of Articulation Rate in Spontaneous Speech. In *Proceedings of Interspeech*, 25–29 August 2013, Lyon, France, 525–529.

Trouvain, J.. 2015. On clicks in German. In A. Leemann, M.-J. Kolly, S. Schmid & V. Dellwo (eds.), *Trends in phonetics and phonology*: *Studies from German-speaking Europe*, 21–34. Bern: Peter Lang.

Trouvain, J., C. Fauth & B. Möbius. 2016. Breath and Non-breath Pauses in Fluent and Disfluent Phases of German and French L1 and L2 Read Speech. In *Proceedings of Speech Prosody (SP8)*, 31 May – 3 June 2016, Boston, USA, 31–35.

Wieling, M., J. Grieve, G. Bouma, J. Fruehwald, J. Coleman & M. Liberman. 2016. Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change* 6(2):199–234.

Winkelmann, R., J. Harrington & K. Jänsch. 2017. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*.

Winkelmann, R., K. Jaensch, S. Cassidy & J. Harrington. 2016. *emuR: Main Package of the EMU Speech Database Management System*.

[i] The example is taken from dialogue multi_C-D_left, 372.2 ms–375.7 ms.

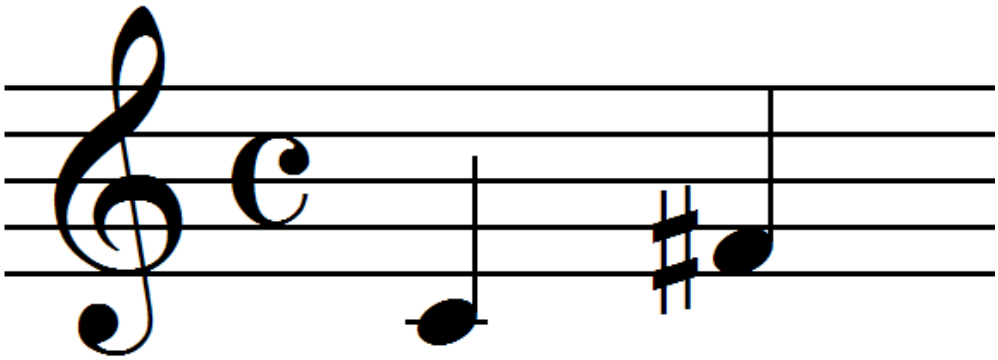
Proceedings of

DiSS 2017

The 8th Workshop on Disfluency in Spontaneous Speech

KTH Royal Institute of Technology
Stockholm, Sweden
18–19 August 2017

TMH-QPSR
Volume 58(1)



Edited by
Robert Eklund & Ralph Rose

Conference website: http://www.diss2017.org
Proceedings also available at: http://roberteklund.info/conferences/diss2017

Cover design by Robert Eklund
Graphics and photographs by Robert Eklund (except ISCA and KTH logotypes)

Proceedings of DiSS 2017, Disfluency in Spontaneous Speech
Workshop held at the Royal Institute of Technology (KTH), Stockholm, Sweden, 18–19 August 2017
TMH-QPSR volume 58(1)
Editors: Robert Eklund & Ralph Rose
Department of Speech, Music and Hearing
Royal Institute of Technology (KTH)
Lindstedtsvägen 24
SE-100 44 Stockholm, Sweden

ISSN 1104-5787
ISRN KTH/CSC/TMH–17/01-SE