

Proceedings of

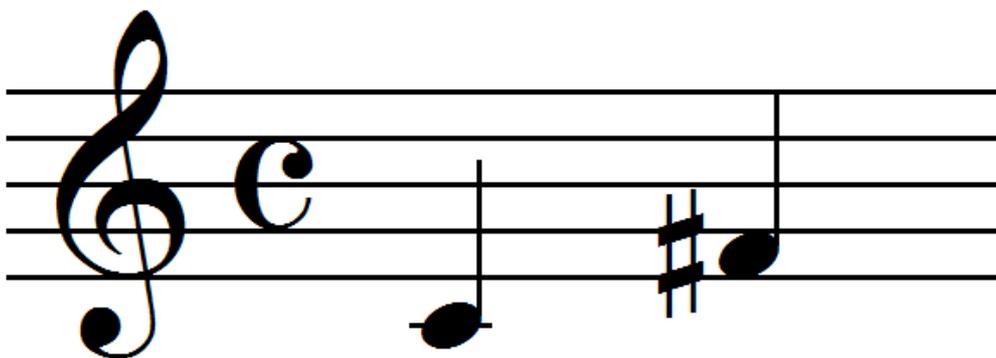


DiSS 2017

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

© The Authors and the Department of Speech, Music and Hearing, KTH, Sweden

Prolongation in German

Simon Betz^{1,2}, Robert Eklund³ and Petra Wagner^{1,2}

¹Phonetics and Phonology Workgroup, Bielefeld University, Bielefeld, Germany

²CITEC, Bielefeld University, Bielefeld, Germany

³Department of Culture and Communication, Linköping University, Sweden

Abstract

We investigate segment prolongation as a means of disfluent hesitation in spontaneous German speech. We describe phonetic and structural features of disfluent prolongation and compare it to data of other languages and to non-disfluent prolongations.

Introduction

We investigate segment prolongation as a means of disfluent hesitation in spontaneous German speech. Prolongation is a common feature of speech occurring near phrase boundaries as a correlate of speakers coming to a halt in articulation. This phenomenon, known as phrase-final lengthening (Turk & Shattuck-Hufnagel, 2007; Umeda, 1977), utterance-final lengthening (Kohler, 1983), prepausal lengthening (O’Shaughnessy, 1995), or boundary-related lengthening (Turk & Shattuck-Hufnagel, 2007) also signals the boundary to the listener (Peters, Kohler & Wesener, 2005).

Prolongation also occurs in disfluent contexts, often in connection with other disfluencies. Within disfluency research there are only a few studies that have dealt with prolongation as a disfluency in its own right, namely corpus studies by Eklund and colleagues (Eklund & Shriberg, 1998; Eklund, 2001, 2004; Den (2003) and Lee et al., 2004) and speech synthesis studies by Betz and Wagner (2016) and Betz et al. (2016, 2017).

In this study, we follow the strand of corpus studies by Eklund and colleagues and present phonetic and structural data on prolongation in German and compare the analyses and distributions to the data available on other languages. In addition, we compare disfluent prolongations to other types of prolongation, showing that there are disfluency-specific features such as syllable position and pitch contour. We use the term *prolongation* with optional extra specifications such as “phrase-final” for all phenomena.

Method and data

We used one part of the DUEL corpus (Hough et al., 2016), called “Dreamapartment”. In this corpus, two speakers have the task to build and furnish the apartment of their dreams in their imagination, with a hypothetical budget of 500.000 € and 200 m² to spare. This results in highly engaged dialogue with frequent disfluency and laughter. Eighteen speakers were recorded in 9 sessions of 30 minutes each,

resulting in 4.5 hours of speech. Speakers were seated next to each other and each speaker was recorded in a separate channel.

The corpus is annotated for disfluencies following an annotation scheme specifically designed for this task (Hough et al., 2015). As there are detection problems regarding prolongations, the corpus has an extra annotation tier for lengthening created semi-automatically (Betz et al., 2017).

In the following section, we present results of the disfluent prolongation corpus data analyses with regard to frequency of occurrence (rates), duration and direct adjacency to other disfluencies, position, morphological complexity, part of speech, segment type and phonological length, and compare it to accentuation lengthening, where appropriate.

Results

Prolongation rates

Prolongation occurrence varies depending on the speaker, between 0.9 and 3.5 per 100 words. On average, there are 1.9 prolonged segments per 100 words with $sd=0.8$. On the time domain, there are 1.6 prolongations per minute of speech (including pauses).

The rate of 1.9% per word is higher than that reported for Swedish (1.27%), Japanese (1.13%; Den, 2003) and American English (0.5%; Eklund & Shriberg, 1998), and lower compared to Mandarin (3.5%; Lee et al., 2004). It has to be considered, however, that comparisons between different kinds of corpora might be difficult, as the DUEL corpus is specifically designed to elicit disfluencies and might thus feature a higher rate of prolongations on average. On the other hand, as shown in Betz et al., (forthcoming), there might be undetected instances of prolongations left in corpora, which would lower the rate accordingly.

Prolongations and fillers

Prolongations are closely linked to other disfluencies. Eklund (2001) reasoned that they might behave similar to fillers as they both signal hesitation by means of vocalization and duration, distinguishing them from other disfluencies, such as silences and repetitions. Adell, Bonafonte and Escudero Mancebo (2008) found in their corpus data that all filled pauses are preceded by prolongations. Betz, Wagner and Voße (2016)

reasoned that this might be related to the phenomenon of phrase final lengthening, as hesitations insert an intonation phrase boundary, which requires prolongation.

Eklund (2001) found that prolongations and filled pauses differ significantly in duration. We can confirm this finding using German data. We compare the phone duration of hesitant prolongation with the duration of prolonged phones preceding a filler and with prolonged phones that a part of a filler. As is shown in Table 1, there is a significant difference in duration. Prolonged phones in fillers are significantly longer than other prolonged phones. Prolongations without contact to fillers are slightly longer than prolongations before fillers, but not significantly. Consequently, Eklund's (2001) conclusion that prolongations and fillers are not similar in function thus receives support from German.

Table 1. Differences in duration.

	t-value(df)	p-value
PR vs. filler	t(63) = -4.6	< 0.001
PR vs. pre-filler	t(40) = 1.96	0.057
Filler vs. pre-filler	t(79) = 5.18	< 0.001
	mean duration (ms)	sd
Prolongation	293.9	130.2
Pre-filler	261.1	78.7
Filler	419.6	19.7

Word position & morphology/syllable structure

In the following, we investigate where hesitant prolongation in German is placed. For illustration, we compare it to prolongation that is due to accentuation from the same dataset.

Swedish is characterized by complex consonant clusters, created by additive affixation of grammatical morphemes, and the maximum allowed complexity of syllables in Swedish is C³VC⁹ (three syllable-initial consonants, and up to nine syllable-final consonants). Given that e.g. Japanese and Tok Pisin are far less permissive in this respect Eklund (2004:251) proposed that PR distribution might be the function of the syllable structure in the language, something Eklund (somewhat misleadingly) called the 'morphology matters hypothesis'. In this respect German is more similar to Swedish, of course.

First, we look at word position, distinguishing three levels: *initial* (first segment in a word), *medial* (a segment in a word that is neither first nor last) and *final* (last segment in a word). There are special cases of one-segmental variants of German words. Most common among these is the indefinite article *ein* which is frequently reduced to *n*. This would be labelled as "final" as the segment it is reduced to originally was word-final according to our definition. As shown in Table 2, disfluent prolongations have a strong tendency to fall on the last segment in a word. In this respect, it differs

from accentuation where medial position is almost as frequent. The observed 7–15–78% distribution found in the word position is markedly different from the observed 30–20–50% distribution reported for American English and Swedish (Eklund & Shriberg, 1998; Eklund, 2001, 2004), two languages with a similar degree of morphological complexity.

Given that the reported distribution in morphologically less complex languages such as Tok Pisin (Eklund, 2001, 2004:251) where the figures are 15–0–85% and Mandarin (Lee et al., 2004), with 4–1–95% and Japanese (Den, 2003), with 0–5–95% suggest that syllable structure plays a vital role in what segment positions are subject to prolongation, but our finding here do not seem to lend support to at least a strong version of Eklund (2004:251).

However, recent findings from Hungarian (Gósy & Eklund, 2017) exhibit a distribution of prolongations similar to that of American English and Swedish, with the figures 18–19–63%. Compared to the very strong tendency found in Japanese and Tok Pisin to produce prolongations mainly on the final segment of words, Hungarian approaches English and Swedish in exhibiting prolongation on initial and medial segments.

Table 2. Word positions for disfluent and accentuation prolongations.

	disfluent	%	accentuation	%
initial	30	7.0	18	19.3
medial	65	15.1	34	36.6
final	336	78.0	41	44.1
Σ	431	100	93	100

Syllable position

We zoom in further and examine syllable position. As can be seen in Table 3, onsets correspond to word-initials and are dispreferred. Most disfluent prolongations are in a syllable's nucleus or coda, whereas accentuation has a strong tendency to fall on the nucleus. This supports the idea that the vocalic core of a syllable is the target for accentuation, whereas a continuant coda is as good for hesitation as a vocalic nucleus.

Table 3. Syllable positions for disfluent and accentuation prolongations.

	disfluent	%	accentuation	%
onset	30	7.0	17	18,2
nucleus	213	49,4	65	70
coda	188	43,6	11	11,8
Σ	431	100	93	100

Segment types, classes and lengths

As summarized in Table 4, sonorants like [m] and [n] outnumber the aggregate of diphthongs and long vowels in being the target of prolongations. Fricatives are more frequent than short vowels. Plosives are very rarely prolonged in German. The instances observed here are either word-initial suspensions of the occlusion (e.g. *das p:asst* (“that fits”) or aspiration added to a word-final stop (e.g. *gut:*h* (“good”). Vowel length is distinctive in German, which is why it makes sense that speakers try to avoid short vowels for hesitant prolongation. Two of the most common words on which disfluent prolongation occurs in German are *und* (“and”) and *dann* (“then”) – both of which have a short vocalic nucleus and both are always prolonged in the final [n] instead of in the nucleus.

The segment type distribution found here exhibits a marked contrast with Swedish, where plosives are frequently prolonged, and where [t] makes the top five list in all corpora examined (Eklund, 2004:247). Once again, Hungarian (Gósy & Eklund, submitted) is similar to Swedish in that all kinds of segments are subject to prolongation.

Open vs. closed class words

Prolongation in German mainly occurs on function words/closed-class words. In the DUEL corpus, this is observed in 62.4% of all cases. In an earlier study on the GECO corpus (Schweitzer & Lewandowski, 2013), the rate is 77% (Betz, Wagner & Voße, 2016).

Table 4. Counts of most frequent phone classes and types. Percentage calculated on the total of 431 instances of disfluent prolongation.

Count	% of total	Phone class
160	37.1	sonorants
150	34.8	diphthongs + long vowels
62	14.4	fricatives
41	9.5	short vowels
10	2.3	plosives
Count		Phone type
98	22.7	n
50	11.6	m
30	7.0	o:
30	7.0	s
22	5.1	ə

While both rates exhibit a strong tendency towards closed-class words, the difference between the two corpora is striking. We can only speculate about the reasons for this. One reason might be the difference in corpus design, GECO being free dialogue and DUEL being highly engaged task-oriented dialogue, which might constrain speaker’s freedom of prolongation placement.

Pitch contour

Research on disfluency pitch exist mainly with regard to fillers (e.g. Adell, Bonafonte & Escudero

Mancebo, 2010; Belz & Reichel, 2015), clitical prolongations that resemble fillers in Japanese (Goto, Itou & Hayamizu, 1999) and Hebrew (Silber-Varod, 2010) or repetitions (Reddy & Hasegawa-Johnson, 2006), but there are no studies on pitch variations of disfluent prolongations.

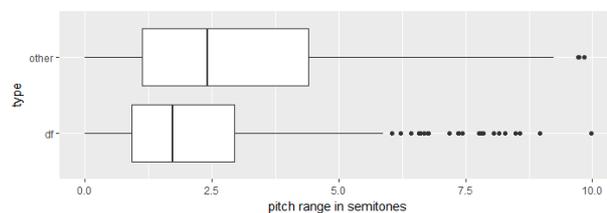


Figure 1. Pitch range differences in disfluent (df) and other prolongations.

We analysed the pitch variations of disfluent and non-disfluent prolongations. The examined data consists of the 431 disfluent prolongations extracted from the DUEL corpus and 250 other prolongations that are not disfluent, but prolonged for other reasons, such as accentuation. Our hypothesis is that pitch is one key feature to distinguish disfluent hesitation prolongations from other types of prolongation, in the sense that hesitations tend to have a flat pitch contour, whereas accentuations naturally exhibit more pitch movement, i.e. pitch accents. To investigate this, we obtained pitch values every 10ms for each instance of prolongation at hand. For each instance, we then calculated the pitch range (in semitones) by subtracting the minimum value from the maximum value. As automatic pitch extraction is known to be prone to errors, we discarded every pitch range value greater than 10 semitones. We then compared the pitch ranges of disfluent and other prolongations using a *t*-test.

As can be seen in Figure 1, non-disfluent prolongations exhibit a higher pitch range with higher variability compared to disfluent prolongations. This difference is significant, $t(543) = 4.07, p < 0.001$. This confirms the hypothesis that there is less pitch movement in disfluent prolongations.

Summary

German exhibits a higher rate of prolongations than most other languages tested in the series of previous corpus studies by Eklund and colleagues, although this might be due to the disfluency-specific design of the corpus at hand.

In terms of duration, German is comparable to Swedish, especially in the sense that fillers are significantly different in duration from prolongations. The preferred segmental targets for prolongations are long vocalic nuclei or sonorants codas. The nuclei will often be word-final, resulting

in a high percentage of word-final prolongations. This is markedly different from Swedish, where consonant prolongation is a common phenomenon, which can also occur word-initially.

In line with earlier studies, we observe a strong tendency for disfluency-related prolongations occurring in closed-class words, although with differences with regard to corpus type.

Pitch variation defines the type of the prolongation: Disfluent prolongations have a comparatively flatter pitch contour compared to other prolongations such as accentuation related ones. For future work, these analyses can be extended to the interaction of prolongations and fillers, for which studies on pitch are available.

Acknowledgements

This research was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

References

- Adell, J., A. Bonafonte & D. Escudero Mancebo. 2010. Modelling filled pauses prosody to synthesise disfluent speech. In *Proceedings of Speech Prosody 2010*, Chicago, USA.
- Adell, J., A. Bonafonte & D. Escudero Mancebo. 2008. On the generation of synthetic disfluent speech: local prosodic modifications caused by the insertion of editing terms. In *Proceedings of Interspeech*, 22–26 September 2008, Brisbane, Australia, 2278–2281.
- Belz, M. & U. D. Reichel. 2015. Pitch characteristics of filled pauses. Presented at *The 7th Workshop on Disfluency in Spontaneous Speech (DiSS)*, 8–9 August 2015, Edinburgh, UK (no page numbers).
- Betz, S., J. Voße, S. Zarriß & P. Wagner. 2017. Increasing recall of lengthening detection via semi-automatic classification. Accepted for *Interspeech 2017*.
- Betz, S., P. Wagner & J. Voße. 2016. Deriving a strategy for synthesizing lengthening disfluencies based on spontaneous conversational speech data. *Phonetik und Phonologie* 12:19–22
- Betz, S. & P. Wagner. 2016. Disfluent Lengthening in Spontaneous Speech. *Studentexte zur Sprachkommunikation (81). Elektronische Sprachsignalverarbeitung (ESSV) 2016*, 135–144.
- Den, Y. 2003. Some strategies in prolonging speech segments in spontaneous Japanese. In R. Eklund (ed.), *Proceedings of DiSS’03, Disfluency in Spontaneous Speech*, 5–8 September 2003, Göteborg, Sweden. *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021, 87–90.
- Eklund, R. 2001. Prolongations: A dark horse in the disfluency stable. In *Proceedings of DISS 2001, Disfluency in Spontaneous Speech*. 29–30 August 2001, Edinburgh, UK, 5–8.
- Eklund, R. 2004. *Disfluency in Swedish human–human and human–machine travel booking dialogues*. PhD thesis, Linköping University, Sweden. ISBN 91-7373-966-9, ISSN 0345-7524
- Eklund, R. & E. Shriberg. 1998. Crosslinguistic Disfluency Modelling: A Comparative Analysis of Swedish and American English Human–Human and Human–Machine Dialogues. In *Proceedings of ICSLP 98*, 30 November – 5 December 1998, Sydney, Australia, 6:2631–2634.
- Gósy, M. & R. Eklund. 2017. Segment Prolongation in Hungarian. In R. Eklund (ed.): *Proceedings of DiSS 2017*, 18–19 June 2017. Royal Institute of Technology, Stockholm, Sweden [this volume], 29–32.
- Goto, M., K. Itou & S. Hayamizu. 1999. A real-time filled pause detection system for spontaneous speech recognition. In *Proceedings of Eurospeech*, 1999, Budapest, Hungary, 227–230.
- Hough, J., L. de Ruiter, S. Betz & D. Schlangen. 2015. Disfluency and laughter annotation in a light-weight dialogue mark-up protocol. Presented at *The 7th Workshop on Disfluency in Spontaneous Speech (DiSS)*, 8–9 August 2015, Edinburgh, UK (no page numbers).
- Hough, J., Y. Tian, L. De Ruiter, S. Betz, D. Schlangen & J. Ginzburg. 2016. DUEL: A Multi-lingual Multimodal Dialogue Corpus for Disfluency, Exclamations and Laughter. In *Proceedings of LREC 2016*, 23–28 May 2016, Portorož, Slovenia, 1784–1788.
- Kohler, K. J. 1983. Prosodic boundary signals in German. *Phonetica* 40(2):89–134.
- Lee, T.-L., Y.-F. He, Y.-J. Huang, S.-C. Tseng & R. Eklund. 2004. Prolongation in spontaneous Mandarin. In *Proceedings of Interspeech 2004*, 4–8 October 2004, Jeju Island, Korea, vol. III, 2181–2184.
- O’Shaughnessy, D. 1995. Timing patterns in fluent and disfluent spontaneous speech. *Proceedings of ICASSP-95*, 9–12 1995, Detroit, Michigan, vol. 1, 600–603.
- Peters, B., K. J. Kohler & T. Wesener. 2005. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K. J. Kohler, F. Kleber und B. Peters (eds.), *Prosodic Structures in German Spontaneous Speech*, Kiel: IPDS, 143–184.
- Reddy, R. M. & M. A. Hasegawa-Johnson. 2006. Analysis of Pitch Contours in Repetition-Disfluency using Stem-ML. *Midwest Computational Linguistics Colloquium*, 2006.
- Schweitzer, A. & N. Lewandowski. 2013. Convergence of articulation rate in spontaneous speech. In *Proceedings of Interspeech 2013*, 25–29 August 2013, Lyon, France, 525–529.
- Silber-Varod, V. 2010. Phonological aspects of hesitation disfluencies. In *Proceedings of Speech Prosody 2010*, 11–14 May 2010, Chicago, USA, 14–19.
- Turk, A. E. & S. Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4):445–472.
- Umeda, N. 1977. Consonant duration in American English. *The Journal of the Acoustical Society of America* 61(3):846–858.