Acoustic characteristics of filler particles in German

Beeke Muhlack[1], Jürgen Trouvain[1], and Michael Jessen[2]
[1]*Language Science and Technology - Saarland University, Germany*
[2]*KT34 Sprache, Audio - Bundeskriminalamt, Germany*

It is generally assumed that filler particles (FPs), such as *äh* and *ähm* in German, are mainly unconsciously produced and thus may prove useful in forensic casework (Jessen, 2008; Künzel, 1987). Disfluencies like FPs, in combination with sound-prolongations, repetitions, and self-interruptions, show a speaker-specific pattern with evidence for German (Braun & Rosin, 2015) and English (McDougall & Duckworth, 2018). However, the consistency of this pattern may be instable across dissimilar speaking tasks, e.g. voice messages compared to interviews (Harrington et al., 2021).

For this study, we aim to present the characteristics of FPs for 100 German male speakers in two conditions: in a Lombard condition and in a non-Lombard ('normal') condition. The data was collected in 2001 as part of the Pool2010 Corpus (Jessen et al., 2005) which uses a picture-description task with forbidden "taboo" words to elicit spontaneous speech. The mean recording time for each speaker is ca. 4 minutes in each condition, amounting to a total duration of ca. 13 h for the sub-corpus investigated here. A typical feature of Lombard speech is an increase of the mean fundamental frequency of speakers as the vocal effort is increased (Jessen et al., 2005). But it is yet unclear to what extent the Lombard condition influences the distribution and phonetic characteristics of FPs. Features under investigation are the frequency (items/min) of different types of FPs (*uh*, *uhm*, *hm*, glottal FPs and tongue clicks), the occurrence of silences before and after the FP, the duration of their segments, fundamental frequency, vowel quality of the vocalic portion of FPs, as well as creaky voice/glottal pulses during the FP.

Preliminary results show that the frequency of typical FPs (*uh*, *uhm*, *hm*) decreases from normal to Lombard speech while the frequency of tongue clicks and glottal FPs (produced with creak/creaky voice only) increases in the Lombard condition (see Table 1). Furthermore, the most frequent FP used by these speakers is the vocalic type (*uh*) which occurs more than twice as often as the vocalic-nasal type (*uhm*). Figure 1 shows that *uhm* is generally longer than *uh*. Moreover, the longest FPs occur between silences, i.e. in a pause. Those FPs that are articulated within an inter-pausal unit (IPU) are shortest. FPs in IPU-final position are longer than in IPU-initial position which is in line with the effect of pre-pausal lengthening. The pattern of a "duration hierarchy" (see Figure 1) holds true for both filler particles *uh* and *uhm*.

Observations on the individual-speaker level reveal that each feature shows high between-speaker variation, so that the rate for the FP *uh* ranges from 0-19 items/min, for *uhm* from 0-15 items/min while *hm* generally shows a lower frequency with a range of 0-7 items/min. The extreme values may be particularly interesting for forensic casework, e.g. three speakers with a higher glottal FP-rate were observed (> 3 standard deviations higher than mean). The poster will show the general trend of the participants but also focus on the individual performance of the speakers and the variation within the dataset.

doi:10.11576/pundp2022-998

	Normal (%)	*Lombard (%)*	**Sum**
uh	921 (36.7)	857 (31.2)	1778
uhm	395 (15.7)	327 (11.9)	722
hm	182 (7.3)	86 (3.1)	268
glottal FP	237 (9.4)	381 (13.9)	618
clicks	774 (30.9)	1098 (39.9)	1872
Sum	2509 (100)	2749 (100)	5258

Table 1. The frequency distribution of the phenomena under investigation in the normal and the Lombard condition. The values in parentheses are the percentages for each condition.

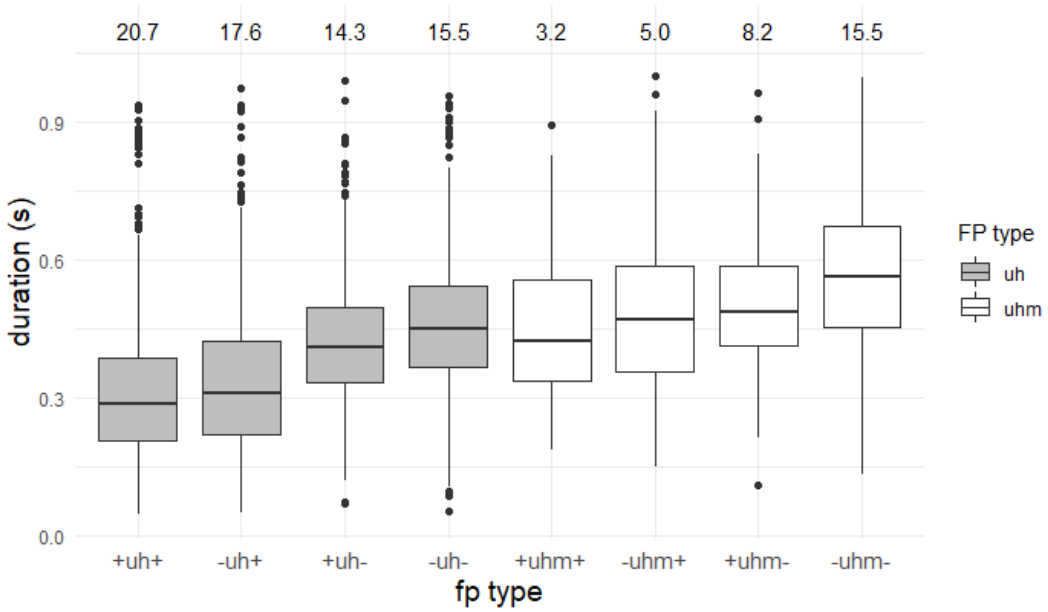

Figure 1. The duration of the filler particles *uh* and *uhm* in their context: speech (+) or silence (-). Thus, a -FP+ occurs in IPU-initial position while +FP- is an FP in IPU-final position, +FP+ occurs within an utterance, -FP- occurs in isolation. The values above each plot represent the percentages per category in the dataset containing only *uh* and *uhm*.

References

Braun, A., & Rosin, A. (2015). On the speaker-specificity of hesitation markers. Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: The University of Glasgow.

Harrington, L., Rhodes, R., & Hughes, V. (2021). Style variability in disfluency analysis for forensic speaker comparison. International Journal of Speech Language and the Law, 28(1), 31–58. https://doi.org/10.1558/ijsll.20214

Jessen, M. (2008). Forensic Phonetics. Language and Linguistics Compass, 4(2), 671–711. https://doi.org/10.1017/S0022226700012755

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. International Journal of Speech Language and the Law, 12(2), 174–213. https://doi.org/10.1558/sll.2005.12.2.174

Künzel, H. J. (1987). Sprechererkennung. Grundzüge forensischer Sprachverarbeitung. Kriminalistik Verlag.

McDougall, K., & Duckworth, M. (2018). Individual patterns of disfluency across speaking styles: A forensic phonetic investigation of Standard Southern British English. International Journal of Speech, Language and the Law, 25(2), 205–230. https://doi.org/10.1558/IJSLL.37241

doi:10.11576/pundp2022-998