ARE THERE INDIVIDUAL DISFLUENCY PATTERNS?

Angelika Braun & Nathalie Elsässer
University of Trier, Germany
brauna @uni-trier.de and s2naelsa@uni-trier.de

ABSTRACT

This contribution takes a look at disfluencies[1] from the forensic practitioner's point of view. It focuses on individuality in the use of disfluency markers. The materials analyzed consist of several minutes of spontaneous speech by eight speakers on three different points in time. Analyses cover fillers including two elements which have not received much attention in previous research: the nasal filler and verbal fillers. Within- and between-speaker differences are assessed. Statistical analysis shows that disfluency markers will distinguish speakers at a level well above chance. At the same time, results show that it is impossible to pin down a single measure which will characterize the disfluency behavior of individual speakers. Rather, a combination of parameters is needed. The forensic implications of these findings are discussed.

Keywords: forensic phonetics, disfluencies, fillers, hesitation

1. INTRODUCTION

Whenever speakers engage in spontaneous conversation, disfluencies, i.e., disruptions of the speech flow are bound to occur. Since about the 1960s, fillers in particular have been looked at primarily as indications of verbal planning and self-monitoring on the part of the speaker (cf. e.g. [18] or [25] [16]) and thus as examples of the symptom function of speech.

The lexical status of fillers has been subject to debate. In this context, Clark and Fox Tree [11] argue vehemently that fillers are normal words (interjections), which would make them signals in the sense of Bühler [9], Corley and Stewart [12], based on an in-depth literature search, summarize that "[t]here is no conclusive evidence that fillers are words" (p. 600).

A key issue in this discussion is to what extent speakers have control over the use of fillers. Clark and Fox Tree [11] find evidence that speakers consciously distinguish between *uh* and *um* in a systematical manner: This is a point which bears relevance to using this parameter in the forensic domain: anything which is subject to active influence on the part of the speaker is prone to voice disguise. If speakers actively control the use of fillers, the value of the parameter for forensic analysis will decrease.

2. INDIVIDUALITY IN DISFLUENCIES

There are observations in various studies pointing to the fact that patterns in the use of disfluency markers may be individual, [19], [16], [2], [18], [10], [13], [21], [25], [1], [11:97-98], [14:6], [15]. This is consistent with the notion that disfluency behavior reflects the cognitive planning process of a specific individual. In the early literature on disfluencies, individuality is stressed much more than in more recent publications. However, none of these studies report data for individual speakers.

A number of issues have been studied in conjunction with disfluencies. Many studies focus on pausing. They generally distinguish between filled and unfilled or silent pauses and among the latter between breath pauses and still pauses [29]. The fillers studied include *uh* and u*m* and their cognates in various languages (e.g. German *äh* and *ähm*, French *euh*; see [1], [11], and [17] for an overview). This, however, does not describe the full range of fillers in actual speech [6].

The present contribution seeks to establish a more comprehensive concept of disfluency than has been done in previous studies. It makes use of the "classical" fillers, which have been studied for decades, but it also proposes new elements which have so far hardly, if ever been considered, such as the nasal filler and verbal fillers. Verbal fillers are multifunctional lexical items [27] which may either carry propositional meaning or serve as fillers. Examples from German are *und* or *ja*, but also phrases which make the search for the appropriate word explicit, such as *wie sagt man* ('how do you put it') or *mir fällt gerade das Wort nicht ein* ('I can't think of the word right now'). In order to shed light on speaker individuality, results are reported separately for each speaker and also by session.

A dedicated forensic approach to disfluencies was developed by McDougall and colleagues [24], [25]. They list a number of parameters which describe the behavioral profile of a given speaker and can be used in forensic casework. This is more comprehensive than any other framework, but it still falls short of being exhaustive, and the intraspeaker consistency of the features is assumed but not tested. The study by Hughes et al. [20] also has a forensic focus. Their approach is confined to studying various aspects of the *um* formant dynamics in a likelihood-ratio based format. They find that to provide relevant information about voice identity, which underlines the importance to include the spectral characteristics

of fillers in the analysis. The key questions to be explored by this research are thus

(a) Are there speaker characteristic features in the disfluency behavior which have so far not been exploited?

(b) Are speakers at all consistent in their disfluency behavior?

(c) Are there features which are suitable for distinguishing between speakers?

3. MATERIALS AND METHODS

The materials used in this study have been analyzed to some degree earlier on [8]. Additional analyses are presented here. Materials consist of recordings from 8 middle-aged female speakers. Subjects' age varied between 45 and 65 years. They were all from the same part of the country. This eliminates gender, age, and dialect as influencing factors. In the present data set, no significant correlation was found between the frequency of disfluencies and speaker age (r = .286; Spearman rank correlation.) Speakers were recorded talking spontaneously about a fixed set of topics (their recent vacation, books they had read, soccer, their opinion on the ban on smoking in restaurants and pubs, what they would tell Angela Merkel if they had a chance to meet her etc.) They were prompted by the investigator once they had nothing more to say on a given topic, but other than that the speech material was monological. Thus, certain factors which have been found to influence disfluency rates, such as relationship, topic, and syntactic complexity [5] could be ruled out for this data set. There were three recording sessions per speaker, which took place about a week apart. This was done in order to be able to look at within-speaker variability vs. between-speaker variability. Sessions lasted between seven and 15 minutes, depending on the frequency of disfluencies and the speaking tempo.

Recordings were analyzed by creating Praat [4] textgrids and annotating the type of pause (breath, silent, filled), the type of filler, the way fillers were worked into the text (preceded and/or followed by a pause), glottal constrictions, verbal fillers, as well as restarts and repairs. The steady phase of the "classical" fillers was marked. It was subsequently used for formant measurements. All annotations were done manually. The following parameters were analyzed by way of Praat scripts: Number of filled and unfilled pauses, number of breath pauses, number and type of fillers, prolongations, connection of fillers with surrounding text, restarts and repairs, formant frequencies of "classical" fillers[2], voice quality of fillers. Only a fraction of results can be presented here.

4. RESULTS

4.1 How to report results

The question of how to report the results is by no means trivial. It has not received much, if any attention in previous studies. Disfluency rates may be expressed in terms of occurrence per lexical item or per time unit. In the English-speaking world, it is common to report disfluency rates per 100 words, cf. e.g. [5], [11], [21]. While this may work for English, it is certainly not a good idea for languages like German which have an almost unlimited potential for compounding, as the word *Bundesausbildungs-förderungsgesetz* ('Federal Education Promotion Act') shows.

The second option is reporting disfluency rates as N per time unit, e.g., per minute. This solves the problem of different word lengths in different languages or speaking styles, but not that of different speech rates. In the forensic context, in which individuality is aimed at, it is crucial to find the measure which best characterizes individual disfluency behavior. In our materials, the correlation between the number of words and the duration of the individual recordings is quite high (r = .63; p = .00093). This means that the results of disfluencies per minute and per 100 words should be similar. This, is indeed the case most of the time, but not if there are large differences in speaking tempo.

4.2 Numbers and proportions

The presentation of results is largely descriptive because the focus is on individual behavior rather than averaged findings [6,7]. Figure 1 shows the proportion of the various disfluency markers per speaker and session.

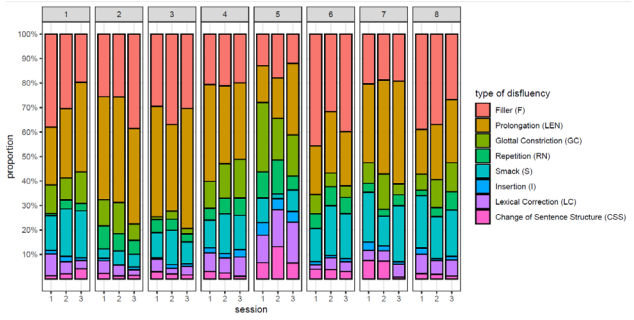

Figure 1: Proportion of disfluency types per speaker and session.

While distributions in the three recording sessions per speaker look strikingly similar, there are notable between-speaker differences. These apply for instance to the relation of fillers (F) to prolongations (LEN). Speakers #3 and #4 show n inverse relationship between the two, whereas #6 stands out by her rare use of prolongations. Other speakers, e.g.,

#7 and #8, are more difficult to distinguish by this parameter.

Table 1: Proportion of vocalic and consonantal prolongations (per 100 words)

prolongations speaker no.	V per 100 words	V %	C per 100 words	C %
#1.1	3.35	54.7	2.77	45.3
#1.2	3.82	61.5	2.45	39.5
#1.3	3.88	45.2	4.71	54.8
#2.1	.58	43.0	.77	57.0
#2.2	.43	19.9	1.73	80.1
#2.3	.78	27.1	2.10	72.9
#3.1	1.78	30.8	4.0	69.2
#3.2	1.79	28.9	4.41	71.1
#3.3	1.29	24.9	3.89	75.1
#4.1	6.88	48.1	7.41	51.9
#4.2	4.67	46.1	5.44	53.9
#4.3	5.54	50.6	5.41	49.4
#5.1	2.72	44.7	3.36	55.3
#5.2	.94	21.1	3,52	78.9
#5.3	1.14	30.3	2,62	69.7
#6.1	.38	15.0	2.15	85.0
#6.2	.53	22.8	1.79	77.2
#6.3	.92	23.1	3.07	76.9
#7.1	.5	16.9	2.45	83.1
#7.2	.95	22.2	3.32	77.8
#7.3	.72	15.4	3.96	84.6
#8.1	1.42	31.2	3.15	68.8
#8.2	.96	20.4	3.75	79.6
#8.3	2.52	39.4	3.88	60.6
Total	**2.09**	**35.07**	**3.87**	**64.93**

Somewhat unexpectedly, prolongations make up the bulk of disfluency markers. For four speakers, their number exceeds that of fillers, three have more fillers than prolongations, and for speaker #1 they are about equal in number. The large proportion of prolongations can possibly be explained by a very meticulous annotation process on our part. Since this was carried out for all speakers alike, the results are comparable. The distribution of vocalic and consonantal prolongations varies between speakers. Table 1 shows the results.

The only study which these results can be compared to is [3]. They report a proportion of vowels of 44.3. The mean proportion of lengthened vowels in our data is 35.1% and thus looks compatible with their results. However, considerable between-speaker differences are evident from the range (15 – 61%). The proportion varies among our speakers but is very consistent over sessions in four of them, whereas two are quite variable. Seven out of eight speakers lengthen consonants more frequently than vowels.

Figure 1 also shows the frequency of occurrence of clicking sounds (S) (smacks). It is evident from the distribution that there are speakers who use this disfluency marker quite frequently. In fact, speaker #7 uses them almost as often as fillers, whereas speakers #3 and #6 hardly exhibit clicks at all.

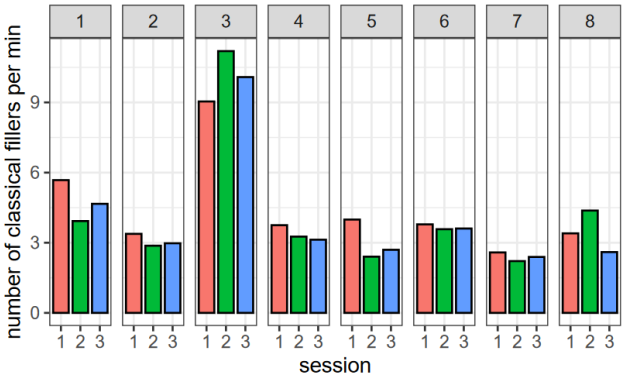

Figure 2: Number of "classical" fillers including *mh* per minute per speaker and session

Figure 2 depicts the number of the "classical" fillers including *mh* per minute. Again, there iss high within-speaker consistency, but this parameter is not suitable to distinguish between speakers very well. Six out of eight speakers are very similar, the remaining two stand out.

The number of fillers per 100 words ranges from 1.3 to 7.3[3]. This considerably exceeds the numbers given by Clark and Fox Tree [11], who find a median of 1.73. One explanation for this difference may be that German, due to its more complex lexical and probably also syntactical structure (the finite verb being at the very end of a subordinate clause), may be more prone to disfluencies than English. If disfluencies per minute are counted, our numbers (2.2 to 11.2) correspond well to those reported by Belz [1] for German in that format, i.e., 1.9 to 11.3.

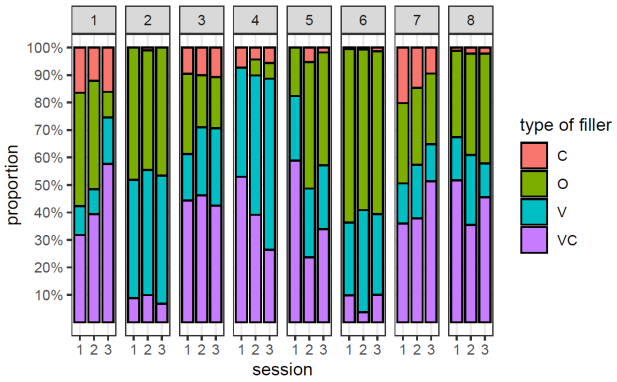

Figure 3: Proportion of types of fillers by speaker and session

Figure 3 depicts the relative proportion of the fillers, i.e,, *äh* (V), *ähm* (VK) and *mh* (K) as well as verbal fillers (termed "O" in the caption) only. As pointed out earlier, the distribution between *äh* and *ähm* is of special interest. Speaker #1, for instance, uses practically no *äh*s, whereas speakers #2 and #6 very rarely exhibit *ähm*. All in all, the majority of subjects (N = 5) show

a clear preference for *ähm,* three for *äh*, and one speaker actually uses verbal fillers more frequently than all others. The preference is clearly speaker specific in our materials. T-tests (unpaired, two-tailed) show a significant preference for *ähm* over *äh* in four speakers (#1: t = -3.82; df = 2.39; p = .046; #3: t = -6.01; df = 2.403; p = .016; #7: t = -4.974; df = 2.566; p = .022; #8: t = -4.314; df = 3.873; p = .013). Two speakers prefer *äh* significantly (#2: t = 26.173; df = 3.916; p = .000; #6: t = 6.09, df = 3.428; p = .005). For the remaining two speakers, results fall short of significance (#4: t = 1.134; df = 3.899; p = .321: #6: t = -1.425; df = 2.011; p = .289). Our data thus do not confirm the observation by Belz [1] and Wieling et al. [30] that speakers, women in particular, generally produce significantly more CV-fillers than V-fillers.

Additional support for considering the nasal filler and verbal fillers to be hesitation markers as opposed to interjections comes from the F0 data. There is no significant difference in this respect between *äh* and *ähm* on the one hand and *mh* and verbal fillers on the other (unpaired T-test, two-tailed; t = 1.064, *p* =.293). This can be taken as yet another indication that these two constitute fillers in their own right. Once again, though, there is one speaker (#7) who does not follow this pattern.

The statistical challenge in this study consists in the fact that in the present context one is trying to prove what is normally considered a confounding factor, i.e., that individual speakers differ. We attempted to do this via a Random Forest model. For training the model, the single takes[4] at all three points in time were treated as separate events. All measurement results in 70% of the takes were used for training. The remaining 30% were used in the classification task. The overall accuracy achieved was 78.5%. This is relatively high considering that the chance level is at 12.5%. The single parameters that were most important for the classification task were expectedly F1, F3, and F4 of the vowel in *äh* and *ähm* as well as, rather unexpectedly, the duration of the inspiration noise.

6. DISCUSSION

Looking at individual disfluency behavior opens a new perspective [7]. In some respects, previous results are confirmed, but it is evident that they can hardly ever be confirmed for all speakers even in this relatively small group. This applies e.g. to filler fundamental frequency. Incidentally, Belz [1] mentions the same thing for his cohort. This is where the forensic interest kicks in. Some findings which have been considered to be well-established can no longer be confirmed if individual speakers are looked at.

As far as the general debate about the lexical status of fillers is concerned, our data do not support Clark and Fox Tree's theory about their being words. One counterargument is the high between-speaker variability paired with a low within-speaker variability of those fillers. In fact, we find the distribution of *äh* vs. *ähm* to be highly individual. If a speaker has a clear preference for one of the two, it can hardly be argued that speakers make a conscious choice in each individual situation.

7. CONCLUSIONS

With the forensic perspective in mind, the following conclusions can be drawn:

(1) There are quite a few parameters in disfluency behavior which would merit being added to the list contained in the TOFFA framework proposed by McDougall et al. [25] along the lines suggested here.

(2) Most results show a striking intra-subject similarity across the three sessions. This applies for instance to the number and type of filler per session, but also the formants. This demonstrates that disfluency behavior is by no means random but instead follows individual patterns, given that the setting is kept constant. That said, while there are speakers who show a large degree of intra-speaker consistency, others are more variable.

(3) In this small group, between-speaker differences exceed within-speaker differences most of the time. Similarity is not a sufficient criterion on which to draw conclusions about speaker identity, though. The typicality criterion also needs to be taken into account [26]. In order to do this, however, much more material needs to be collected. Typicality will have to be assessed with respect to a speaker pattern as opposed to singular disfluency markers.

It is by no means the intention of this research to argue that speaker identification should rely on disfluency behavior alone. But the present results do indicate that while no single parameter is likely to suffice to distinguish between speakers, looking at the complete disfluency pattern has the potential of doing just that.

It might be argued that this approach is unrealistic in the sense that recordings of sufficient duration will hardly be available in a forensic setting. That is only partly true, though. Particularly in jurisdictions which allow telephone intercepts, there are often many minutes of both questioned and reference speech material available from the same telephone surveillance measure. A detailed analysis of the disfluency behavior is therefore possible.

ACKNOWLEDGMENTS

A warm word of thanks goes to Melissa Hildebrand and Vivien Meyer for doing much of the tagging.

7. REFERENCES

[1] Belz, M. 2021. Die Phonetik von *äh* und *ähm*. Akustische Variation von Füllpartikeln im Deutschen. In: Eklund, R. (ed.) *DiSS. The 7th Workshop on Disfluency in Spontaneous Speech*.. Berlin, 1–4.

[2] Belz, M. and J. Trouvain. 2019. Are 'silent' pauses always silent? *Proc. 19th ICPhS,* Melbourne.

[3] Betz, S., Eklund, R., and Wagner, P. 2017. Prolongation in German. In *DiSS 2017 The 8th Workshop on Disfluency in Spontaneous Speech, KTH, Royal Institute of Technology, Stockholm, Sweden,* pp.13-16.

[4] Boersma, P., Weenink, D. 2022. *Praat: doing phonetics by computer* [Computer program]. Version 6.3.03, retrieved 17 December 2022.

[5] Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., Brennan, S.E. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech* 44(2), 123-147.

[6] Braun, A. 2021 Forensische Sprach- und Signalverarbeitung. In J. Bockemühl (ed.) *Handbuch des Fachanwalts Strafrecht*. 8th ed. Köln: Carl Heymanns Verlag, 1890 – 1914.

[7] Braun, A. 2021: Nonverbal vocalizations – A forensic phonetic perspective. *Proceedings Workshop on Laughter and other Nonverbal Vocalisations,* Bielefeld 2020.

[8] Braun, A., Rosin, A. 2015. On the speaker specificity of hesitation markers – a pilot study. *Proceedings of ICPhS Glasgow,* 5p.

[9] Bühler, K. 1934. *Sprachtheorie. Die Darstellungs-funktion der Sprache*. Jena: G. Fischer.

[10] Butcher, A. 1973. *Aspects of the perception and production of pauses in speech.* Arbeitsberichte of the Phonetics Institute of the University of Kiel, Nr. 1.

[11] Clark, H. H., Fox Tree, J.E. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*(1), 73-111.

[12] Corley, M., Stewart, O. M. 2008. Hesitation Disfluencies in Spontaneous Speech: The Meaning of *um*. *Language and Linguistics Compass* 2/4, 589

[13] Duez, D. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech* 25, 11-28.

[14] Eklund, R. 2001. Prolongations: A dark hoarse in the disfluency stable. In *Proceedings of DISS '01*, Edinburgh, 5-8.

[15] Fant, G., Kruckenberg, A. and Barbosa Ferreira, J. 2003. Individual variation in pausing. A study in read speech. *PHONUM* 9, 193-196.

[16] Finlayson, I.R, Corley, M. 2012. Disfluency in Dialogue: An Intentional Signal from the Speaker? *Psychon Bull Rev* 19, 921-928.

[17] Fuchs, S., König, L. L., Gerstenberg, A. 2021. A Longitudinal Study of Speech Acoustics in older French Females: Analysis of the Filler Particle *euh* across Utterance Positions. *Languages*, *6*(4): 211; https://doi.org/10.3390/languages6040211 .

[18] Goldman-Eisler, F. 1961. A comparative study of two hesitation phenomena. *Language and Speech* 4: 18-26.

[19] Goldman Eisler, F. 1968. *Psycholinguistics. Experiments in Spontaneous Speech*. London and New York: Academic Press.

[20] Hughes, V., Foulkes, P., Wood, S. 2016. Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law*, 23(1): 99-132.

[21] Kjellmer, G. 2003. Hesitation. In defence of ER and ERM. E*nglish Studies* 84(2): 170–198. Open yale courses. https://oyc.yale.edu/. Accessed: June 6th, 2022.

[22] Levelt, W J.M. 1983.Monitoring and self-repair in speech. *Cognition* 14: 41-104.

[23] Maclay, H., Osgood, Ch.E. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15: 19-44.

[24] 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

[25] McDougall, K., Rhodes, R., Duckworth, M., French, P., Kirchhübel, Chr. 2019. Application of the 'Toffa' Framework to the Analysis of Disfluencies in Forensic Phonetic Casework. In: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds), *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc, pp. 731-735.

[26] Rose, Ph. *Forensic Speaker Identification*. London/New York: Taylor and Francis.

[27] Stenström, A-B. 2012. Pauses and hesitations. In G. Andersen and K.Aijmer (eds.), *Pragmatics of Society.* Berlin/Boston: De Gruyter Mouton, 537-567.

[28] Trouvain, J., Fauth, C., Möbius, B. 2016. Breath and Non-breath Pauses in Fluent and Disfluent Phases of German and French L1 and L2 Read Speech. *Proceedings of Speech Prosody (SP8)*, 31-35.

[29] Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., Liberman, M. 2016.: Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change* 6 (2): 199–234. doi:10.1163/22105832-00602001.

[1] *Disfluency* is used as a cover term for hesitations (covert repairs) including fillers and (overt) repairs [22].

[2] Four formants were measured.

[3] The corresponding graph is not given here for reasons of space. Both had to be calculated since reference units differ in the literature.

[4] One „take" is defined as a part of a recording dealing with one topic.