

ON THE SPEAKER-SPECIFICITY OF HESITATION MARKERS

Angelika Braun and Annabelle Rosin
University of Trier, Germany
brauna@uni-trier.de; annabelle.rosin@gmx.de

ABSTRACT

The occurrence of hesitation markers is generally considered to be part of the verbal planning process. It is also a feature which is of potential importance to the forensic application of phonetics if hesitation behaviour could be linked to individual speakers. This study examines a total of eight female speakers on three different days.

It can be demonstrated that, even though results vary across sessions, subjects exhibit distinct patterns of hesitation marker usage. This pertains to the number as well as the type of hesitations marker, which makes this feature a potential candidate for forensic investigations.

Keywords: verbal planning, forensic phonetics

1. INTRODUCTION

Whenever speakers engage in spontaneous conversation, disfluencies, i.e., disruptions of the speech flow are bound to occur. These disfluencies are generally regarded as indications of verbal planning and/or monitoring of the speech signal (cf. e.g. Goldman-Eisler [14] or Shriberg [22]). Phonetic manifestations are varied – they range from pausing to hesitation markers of various sorts and false starts or repetitions.

Earlier studies have addressed the prosodic relationship between hesitation markers and the surrounding speech utterances. Specifically, the question was how hesitation markers were worked into the intonational pattern of the respective sentence. The fundamental frequency of these sounds as compared to the surrounding words has been studied for a variety of languages: American and British English (Shriberg [21, 23]), German (Batliner et al. [2]), Dutch (Swerts [25]), Spanish (Adell et al. [1]), French (Duez [10, 11]), Mandarin (Zhao [26]), and Arabic, Italian and Portuguese (Candea et al. [8]). Numerous researchers found a lowering of the fundamental frequency of hesitation markers relative to their immediate context (cf. Shriberg / Lickley

[23], Batliner et al. [2], Swerts [25], Bauer [3], Candea et al. [8], Zhao / Jurafsky [26]).

This contribution focusses on hesitation markers, i.e., sounds which speakers produce as a function of the verbal planning process. These encompass filled pauses as well as initial and, predominantly, final lengthening. The principal question beyond fundamental frequency issues is to what extent these hesitation markers are speaker-specific. This is of particular interest to neurolinguists, because if individual patterns were to be established, this would point to individual processing strategies. In this context, it is also of interest to forensic phoneticians who seek to establish any incidence of speaker-specificity in the speech material.

A number of studies point to the fact that speakers exhibit individual patterns of hesitaion marker usage (Maclay / Osgood [19], Goldman-Eisler [13], Blankenship / Kay [4], Henderson / Goldman-Eisler [16], Goldman-Eisler [14], Butcher [7], Duez [9], Kowal [17], Shriberg [22], Olbertz-Siitonen [20]). This is consistent with the notion that hesitation behavior can hardly be consciously influenced by the speaker but instead reflects the cognitive planning process of a specific individual. The speaker-specificity of hesitation marker use can be regarded as a starting point for the present study.

The following research questions are addressed

- Is there any intraspeaker consistency across time with respect to both type of hesitation sound and fundamental frequency thereof?
- Is hesitation behavior speaker-specific?
- Is the fundamental frequency of hesitation markers lower relative to the immediate context as well as to the overall mean of the respective speaker?

2. MATERIALS AND METHODS

A total of 10 female speakers¹ were recorded on three different days. They were between 45 and 65 years old with a mean of 56.4. All participants speak German as their first language and are self-reported

¹ Since gender-effects on hesitation behavior cannot be ruled out and the number of subjects had to be kept relatively low, it was decided to use female speakers only.

right-handers. The latter aspect is well worth taking into account because stuttering as well as dyslexia are frequently cited as a consequence of inconsistent hemispheric lateralization (Böhme [5], 124; Eglington / Annett [12]). Subjects were chosen who speak the same regional variety of German (Hessian) in order to avoid any effects which might be attributed to regional accent. Participants were questioned about their awareness state at the time of the recording – none reported any extraordinary fatigue which could have influenced their hesitation behavior.

The recordings were carried out at the subjects' homes in order to ensure that they felt at ease with the situation. A digital recorder Zoom H4 was used for the recordings. The sampling rate was set to 44.1 KHz. Speakers were asked to produce a few minutes of spontaneous speech on a topic of their choice. If they could not think of anything (further) to talk about, they were prompted by a sequence of words which were to be worked into a story. Initially, a minimum of 100 hesitations per person per session was aimed at, but this could not always be achieved. Two subjects had to be excluded in the process because they either did not produce nearly as many hesitations as were expected or did not complete the three recording sessions. Thus, a total of 8 subjects who had been recorded for a total of 44 to 75 minutes in total remained for analysis.

The recordings were transcribed, and the following types of hesitation markers were identified:

- fillers of various sorts (vowel, vowel + nasal, nasal)
- initial vowel lengthening
- initial consonant lengthening
- final vowel lengthening
- final consonant lengthening

All other hesitation phenomena (pauses, repetitions, false starts) were disregarded.

Fundamental frequency measurements were carried out across the full length of the samples as well as for fillers and their immediate context (+/- 2 seconds) only. The voice report option of the PRAAT software package (Boersma / Weenink [6]) was used.

3. RESULTS

3.1 Quantitative analysis

The quantitative analysis of the hesitation markers shows the following results:

Ss	1	2	4	5	6	7	8	10	Total
#1	77	58	154	99	78	142	109	118	835
#2	83	85	227	113	64	124	134	154	984
#3	161	87	172	99	55	172	122	125	993
Total	321	230	553	311	197	438	365	397	2812

Table 1: Number of hesitations per speaker and session

This table indicates that some speakers may use more hesitation markers than others. Even though 100 hesitations per session would have been desirable, speakers no. 2 and 6 clearly fell short of this goal. The range in individual hesitation frequencies becomes even more evident if the number of hesitations per time unit is considered (cf. Table 2).

Ss	1	2	4	5	6	7	8	10
#1	8.6	4.8	11.0	5.5	5.6	7.1	4.5	6.2
#2	5.5	4.3	13.3	5.1	3.8	6.2	5.0	7.7
#3	7.0	4.6	12.3	6.6	4.2	7.2	5.1	6.3
Total	6.8	4.5	12.3	5.7	4.5	6.8	4.9	6.7

Table 2: Number of hesitations per speaker per minute

The number of hesitations per minute varies greatly between speakers, but can certainly not serve to reliably discriminate between them. For instance, speakers no. 2, 6, and 8 show average frequencies of hesitating which are very similar, but they also exhibit different degrees of intra-speaker variability. It is conceivable that the frequency of hesitation phenomena may be related to speaking tempo which in turn has been shown to be speaker specific by some authors (cf. Grosjean and Deschamps [15], 154f. and Goldman Eisler [14], 19), but this is not followed up in this contribution.

In addition to the number of hesitation markers per time unit, it is worth looking at their internal distribution, i.e., the frequency of their phonetic manifestations. As mentioned earlier, seven different types of hesitation markers were studied: insertions of a vowel (“uh”), a nasal (“mh”), or a sequence of

vowel + nasal (“um”); initial vowel or consonant lengthening; final vowel or consonant lengthening. Figure 1 shows the results for all speakers and sessions.

This figure demonstrates a number of crucial findings: In the first place it shows that not every speaker makes use of all seven options. Instead, speakers generally use four to five different hesitation markers with a focus on one to two. Whereas there is no one “favorite” hesitation marker which is preferred by a majority of speakers, there are two which are very rarely used: initial vowel lengthening and initial consonant lengthening.

Subjects’ individual patterns vary across sessions, but not to an extent where they would easily be confused with any of the other speaker’s patterns. Three speakers whose frequency of hesitation marker use was very similar (nos. 2, 6, and 8) can now easily be discriminated based on the distribution of hesitation sounds. While speaker no. 2 shows a preference for the "uh" type hesitation, no. 6 clearly prefers "um". No. 8, finally, exhibits a clear preference for final lengthening. At first glance, speaker no. 2 and speaker no. 7 resemble each other, but a closer look reveals that speaker no. 2 is much more consistent than speaker no. 7, who also used a wider range of hesitation markers. Incidentally, those two speaker would never have been confused since their frequency of hesitating is very different. Still, the pattern of fillers alone will probably not suffice to differentiate between speakers. This is why their fundamental frequency relative to the immediate context and the overall mean is also considered.

3.2 Fundamental frequency

3.2.1 Hesitations and overall mean

The mean fundamental frequency of the full recordings was measured for all subjects and all sessions. Table 4 shows the results.

The results are somewhat unexpected in the sense that subjects no. 2, 7, and 8 show average fundamental frequencies which are unusually low for female speakers as compared to normative data for German (Künzel [18], 83). This may in part be due to a general tendency for female speakers in Europe to emphasize the lower end of their modal register as part of their gender identity. In addition, most subjects studied here were either menopausal or post-menopausal, and thus their hormone status can be expected to have lowered their fundamental frequency (cf. Stoicheff [24]). At any rate, in the context of the present study only the relationship between the mean fundamental frequency and that of the hesitation markers is of interest.

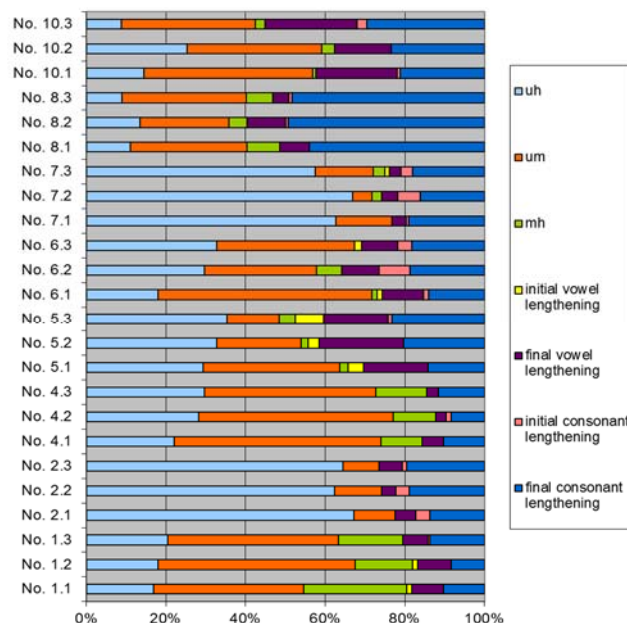


Figure 1: Phonetic manifestations of hesitation markers according to speakers and sessions

Subjects	1	2	4	5	6	7	8	10
Session 1								
Mean F0	184	168	199	184	209	158	165	199
Session 2								
Mean F0	182	158	199	188	212	163	178	193
Session 3								
Mean F0	190	158	192	190	212	167	174	208

Table 3: Mean F0 of all subjects in all sessions

In order to carry out fundamental frequency measurements of the hesitation markers, all fillers per speaker were concatenated into one file, and subsequently their average F0 was determined. Thus, a total of 1339 fillers were measured. This number is notably smaller than the total number of hesitations which are contained in the corpus studied here. This is due to the fact that lengthenings were not considered at all in this context and repeated hesitations of the type “uh, mh, um” were counted as one incident each. Table 4 shows the mean fundamental frequency for the accumulated hesitation sounds.

Subjects	F0-text	F0-hes
1	185	176
2	161	159
4	197	180
5	187	164
6	211	194
7	163	159
8	172	175
10	200	183

Table 4: Mean F0 of the total utterance (F0-text) averaged over sessions and mean F0 of the concatenated hesitations (F0-hes)

Seven out of the eight speakers show a (marked) decrease in F0 for the hesitation sounds as compared to the average fundamental frequency. Only speaker no. 8 exhibits a small increase of 3 Hz for the hesitations. A difference of this magnitude, however, is hardly perceptible. Statistical analysis (Wilcoxon test) shows the difference between means to be significant ($p = .012$).

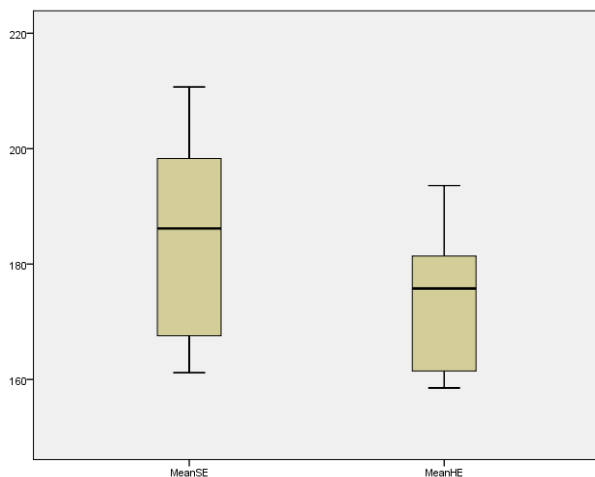


Figure 2: Box plot of F0 means for text (left) and hesitations (right)

4. DISCUSSION

Overall, the results of the present study confirm previous findings. Generally, the F0 of hesitations showed results which were to be expected, i.e., a lowering of mean F0. Speakers demonstrated a pattern of hesitation phenomena which was fairly consistent across recordings. At the same time, a speaker-specific frequency of occurrence of hesitations was observed. Furthermore, indications of speaker-specific distributions among the hesitation

markers were found. This confirms findings by Maclay / Osgood [19], Blankenship / Kay [4], Goldman Eisler [14], Shriberg [22] and Olbertz-Siitonen [20].

As far as hesitation marker F0 as compared to the average F0 of a given speaker and the immediate context is concerned, highly ($p = .012$) and very highly (.000) significant differences were found, respectively. The F0 of hesitation markers was found to be generally lower than in the other conditions.

This confirms the results cited in studies by Shriberg / Lickley [23], Batliner et al. [2], Duez [10, 11], Adell et al. [1]. However, this finding does not seem to be universal – hesitation markers of the type “mh” in particular seem to be prone to be uttered at a higher F0 than the surrounding utterances.

5. CONCLUSIONS

This small-scale pilot study demonstrates that the number of hesitation markers as well as their distribution are promising features of speaker identity. The search for individual patterns of hesitation behavior can be carried even further. For example, the positioning within the utterance may result in additional differentiation between speakers. Also, the formant structure of fillers as well as the way they are worked into the utterance (connected to preceding or following word vs. surrounded by pauses) ought to be taken into account. On the other hand, it is evident that more material is needed to reliably judge a given speaker’s hesitation behavior than is available in many forensic cases. That is why hesitation sounds will hardly lend themselves to being used on their own, but in combination with other well-established speaker-specific features they may serve to support the expert’s judgement.

6. REFERENCES

- [1] Adell, J., Bonafonte, A., Escudero, D. 2006. Disfluent Speech *Analysis and Synthesis: a preliminary approach*. Proc. 3th International Conference on Speech Prosody, Dresden, Germany, n.pag.
- [2] Batliner, A. et al. 1995. Filled pauses in spontaneous speech. Proc. XIIIth ICPPhS, Stockholm, Vol. 3, 472-475.
- [3] Bauer, D. 2005. Gefüllte Pausen als Persönlichkeitsmarker in der forensischen Phonetik, Abstract. <http://38.stuts.de/indexb3de.html?option=com>

- _content&task= view&id=42&Itemid=78, Stand: 11.04.2011.
- [4] Blankenship, J., Kay, Ch. 1964. Hesitation Phenomena in English Speech: A Study in Distribution. *Word* 20, 360-372.
- [5] Böhme, G. (ed.) 2003. *Sprach-, Sprech-, Stimm- und Schluckstörungen. Bd. 1: Klinik*. 4., aktualisierte und erweiterte Auflage. München / Jena: Urban & Fischer.
- [6] Boersma, P., Weenink, 2013. PRAAT: doing phonetics by computer. Available from <http://www.fon.hum.uva.nl/praat/>.
- [7] Butcher, A. 1973. *Aspects of the perception and production of pauses in speech*. (English translation of extracts from Arbeitsberichte Nr. 1 of the Phonetics Institute of the University of Kiel, 1973).
- [8] Candea, M., Vasilescu, Ioana, Adda-Decker, Martine 2005. Inter- and intra-language acoustic analysis of autonomous fillers. *Proc. DiSS'05, Disfluency in Spontaneous Speech Workshop*, Aix-en-Provence, France, 47-51.
- [9] Duez, D. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech* 25, 11-28.
- [10] Duez, D. 2001a. Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français. *Travaux Interdisciplinaires du Laboratoire Parole et Langage* 20, 31-48.
- [11] Duez, D. 2001b. Acoustico-phonetic Characteristics of Filled Pauses in Spontaneous French Speech: Preliminary Results. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech (DiSS'01)*, Edinburgh, Scotland, UK, 41-44.
- [12] Eglinton, E., Annett, M. 1994. Handedness and dyslexia: a meta-analysis. *Perceptual and Motor Skills* 79, 1611-1616.
- [13] Goldman-Eisler, F. 1961. A comparative study of two hesitation phenomena. *Language and Speech* 4, 18-26.
- [14] Goldman Eisler, Frieda 1968. *Psycholinguistics. Experiments in Spontaneous Speech*. London and New York: Academic Press.
- [15] Grosjean, F., Deschamps, A. 1975. Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31, 144-184.
- [16] Henderson, A., Goldman-Eisler, F., Skarbak, A. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech* 9, 207-216.
- [17] Kowal, S. 1991: *Über die zeitliche Organisation des Sprechens in der Öffentlichkeit. Pausen, Sprechtempo und Verzögerungen in Interviews und Reden von Politikern*. Bern, Stuttgart, Toronto: Verlag Hans Huber.
- [18] Künzel, H. J. 1987. *Sprechererkennung. Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.
- [19] Maclay, H., Osgood, Charles E. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15, 19-44.
- [20] Olbertz-Siitonen, M. 2007.: ähm vs. niinku – Verzögerungssignale in deutschen und finnischen Diskussionen. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 12 (2), 1-22.
- [21] Shriberg, E. 1995. Acoustic properties of disfluent repetitions. *Proc. International Congress of Phonetic Sciences*, 4, Stockholm, Sweden, S. 384-387.
- [22] Shriberg, E. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31, 153-169.
- [23] Shriberg, E. E. / Lickley, R. J. 1993. Intonation of Clause-Internal Filled Pauses. *Phonetica* 50, 172-179.
- [24] Stoicheff, M. L. 1981. Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech and Hearing Research* 24, 437-441.
- [25] Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485-496.
- [26] Zhao, Y. Jurafsky, D. 2005. A preliminary study of Mandarin filled pauses. *Proc. DiSS'05, Disfluency in Spontaneous Speech Workshop*, Aix-en-Provence, France, 179-182.