

The following paper was presented at The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019) held at ELTE Eötvös Loránd University in Budapest, Hungary on 12–13 September, 2019.

Title: Variation in the choice of filled pause: A language change, or a variation in meaning?

Author(s): Hong Zhang

Abstract: The role of filled pauses in message structuring is a heavily debated question, but the result is still somewhat inconclusive. In this study, I consider this question jointly with sociolinguistic factors that have been thought to affect the choice of filled pause in American English. The results suggest that the use of *uh* is subject to higher variability across not only age groups, but also conversation topics and interlocutors. A latent semantic analysis found consistent difference between two forms of filled pause and silent pauses of varying duration in the primary latent dimension, but similarity between short silent pause and *uh*, as well as long silent pause and *um* in the second dimension. Therefore, the functional difference between *um* and *uh* should be acknowledged, and the observed change in their relative popularity is potentially related to their different meaning or function in the discourse.

DOI: <https://doi.org/10.21862/diss-09-005-zhang>

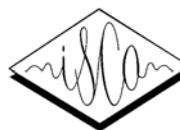
Citation (JIPA): Zhang, Hong. 2019. Variation in the choice of filled pause: A language change, or a variation in meaning? In: R. L. Rose & R. Eklund (eds.), *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, 12–13 September, 2019, Budapest, Hungary, 15–18.

The complete proceedings for DiSS 2019 are available as follows.

ISBN: 978-963-489-063-8

DOI: <https://doi.org/10.21862/diss-09>

DiSS 2019 was sponsored by The Faculty of Humanities, ELTE Eötvös Loránd University and the International Speech Communication Association (ISCA).



# Variation in the choice of filled pause: A language change, or a variation in meaning?

Hong Zhang

Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA

## Abstract

*The role of filled pauses in message structuring is a heavily debated question, but the result is still somewhat inconclusive. In this study, I consider this question jointly with sociolinguistic factors that have been thought to affect the choice of filled pause in American English. The results suggest that the use of uh is subject to higher variability across not only age groups, but also conversation topics and interlocutors. A latent semantic analysis found consistent difference between two forms of filled pause and silent pauses of varying duration in the primary latent dimension, but similarity between short silent pause and uh, as well as long silent pause and um in the second dimension. Therefore, the functional difference between um and uh should be acknowledged, and the observed change in their relative popularity is potentially related to their different meaning or function in the discourse.*

## Introduction

Two forms of filled pause in American English, *um* and *uh*, have long been the focus of studies on speech disfluencies. In addition to understanding their relation to the cognitive process of speech production (e.g. Goldman-Eisler, 1968; Rochester, 1973; Levelt, 1983), debates have also been centered around whether the two fillers play different roles in structuring the message. Clark and Fox Tree (2002) argue that the fillers are different words which carry distinctive discourse meanings, citing evidence from the different delay after the fillers and the different role in facilitating word recognition. However, a later study (O'Connell & Kowal, 2005) failed to find the functional difference through analyzing media interviews of Hillary Clinton. In a review paper by Corley and Stewart (2008), they concluded that there is not clear evidence that speakers have intentional control over the choice of fillers.

More recently, correlations between sociolinguistic factors, such as age, gender and socioeconomic status of the speaker, and variation in the choice of filler forms have been thoroughly examined both in English (Acton, 2011; Tottie, 2011; Fruehwald, 2016) and several other Germanic languages (Wieling et al., 2016). Among the sociolinguistic factors, an interesting observation is

that the relative frequency of *um* over *uh* is higher in younger speakers, where female speakers seem to lead the trend. Therefore, it has been proposed that the choice of filled pause is in fact a sociolinguistic variable and a language change in progress in the Germanic family.

However, two potential covariates which in principle partially explain the variation in the choice of filled pause: the topic of conversations and the accommodation between interlocutors, have not been explicitly and systematically discussed in the sociolinguistic literature. In this study, I show that conversation topic and the effect of interlocutor are able to reveal a possible functional difference between the two filled pauses, and this difference may also point to valid explanations of the proposed change in progress. However, this difference may or may not be the kind as proposed in Clark and Fox Tree (2002), and should only be understood in relation with other potential influencing variables.

## Data and method

### The corpus

The data used for this study is from the Fisher Corpus of American English (Cieri, Miller & Walker, 2004). This corpus contains 16,454 10-minute telephone conversations recorded in separate channels, totaling 2,742 hours of speech. Conversations were guided by 40 topics. A subset of 9,471 one-sided speech from 3,157 native speakers of American English are selected to form the analysis sample, representing about 790 hours of spontaneous conversations. The selected speakers all contributed in exactly three conversations.

### Data pre-processing

The analysis sample is forced aligned with Penn Forced Aligner (Yuan & Liberman, 2008). Each conversation was first segmented into turns based on the time stamps and speaker identification number provided in the transcription before passing to the forced aligner. Back channel talking and floor holding were also removed by filtering out turns shorter than four words and containing only filler words. Durations were obtained from forced alignment results.

## Sociolinguistic factors

### Age and gender

Age and gender's effect on the alternation between two forms of filled pause is most profoundly discussed in sociolinguistic literature. The most noticeable change, according to Fruehwald (2016) and Wieling et al. (2016), is that *um* is trading frequency with *uh* among younger speakers. This reported trend serves as the main evidence for the "language change in progress" argument (Labov, 1994).

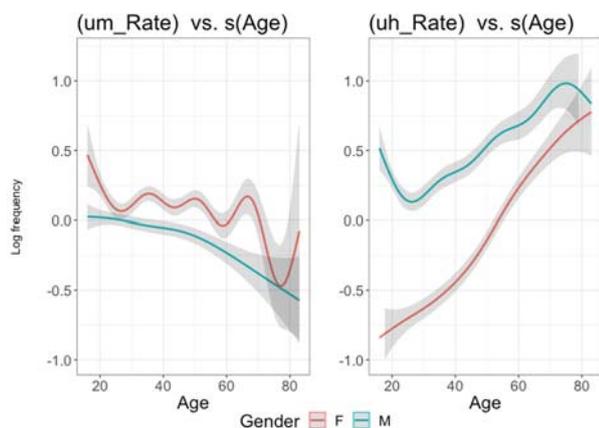


Figure 1. Age and gender effect on the frequency of filled pause by form.

Figure 1 plots the per-speaker frequency of the two fillers as the function of age, grouped by gender. Regression lines are fitted through a quasi-Poisson regression with per-speaker as the response, and shaded areas represent 95% confidence bands. The figure clearly shows that the change in frequency is mainly driven by the change in *uh*. On the other hand, the frequency of *um* is largely stable for female speakers between 20 and 60 years of age, and only a slight decrease can be found among male speakers. Inspecting the effect of age on the overall per speaker frequency of filled pause, as shown in Figure 2, it is observed that there is an overall increase in the frequency as people getting older. Male speakers consistently have higher frequency than female speakers, just as in the case of *uh*. Therefore *um* appears to be time-invariant, while *uh* displays more change as a function of age. Therefore the term 'trading frequency' may be an exaggeration.

### Individual variation

The individual variation of *um/uh* choice can be effectively visualized through the 2-D density plot with the frequency of *um* and *uh* as two dimensions, as plotted in Figure 3.

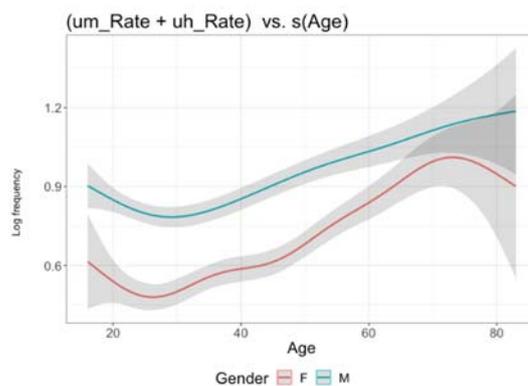


Figure 2. Age and gender effect on the overall frequency of filled pause.

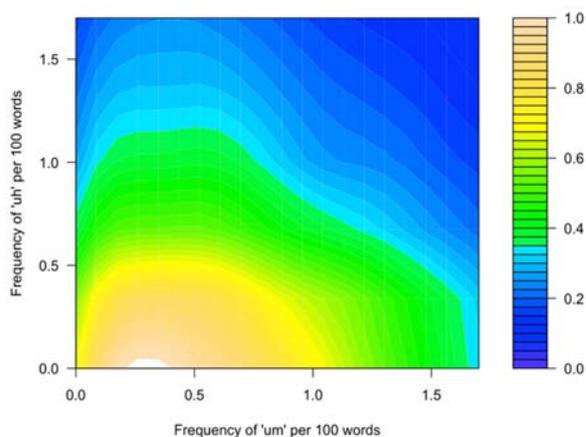


Figure 3. Individual variation in the choice of filled pause plotted as per speaker frequency of *um* and *uh*.

Two noticeable patterns can be observed from Figure 3. First, there is an inverse correlation between the frequency of *um* and *uh*, but this relation only holds for speakers who use more *um*. Second, the highest density is along the *um* axis, meaning that there are more predominantly *um* users. Thus for a given speaker, there does seem to be a preference for one filler over the other, and *um* appears to be the more popular choice.

### Influence of topic and interlocutor

The effect of topic on individual choice of fillers can be visualized on the same dimensions shown in Figure 3. The median frequencies across topics are plotted in Figure 4. It is obvious that greater variation can be found along the *uh* frequency dimension. This pattern offers another piece of evidence that *uh* has greater variability compared to *um* across different conversation topics. Combining the observations from Figure 3, one plausible hypothesis is that for many people, *um* is the default choice of filled pause for the need to insert a filler, while *uh* comes in more involuntarily as a response to other challenges in

speech production (i.e. they have different meanings).

This and other similar hypotheses can be supported by two further observations: The patterns of between-topic and between-speaker correlations of the frequency of fillers. Figure 5 is the quantile plot of between-topic correlations of *um* and *uh* frequencies. The pairwise correlations between topics are not only higher for *um* for about half of the topic pairs, the values of correlation coefficients are also almost consistently greater than 0.9. On the contrary, about 40% of the correlations are less than 0.9 for *uh* frequency, and the lowest correlation is lower than 0.6. Considering the inventory of topics in Fisher, the variation among *uh* frequencies, as well as the contrast between *um* and *uh*, is non-trivial. To give an example, the topics provided to speakers ranged from Iraqi War to Food and Personal Hobbies. The split between more serious social or political topics, and topics on more casual or hypothetical situations is about 3 to 8. More similarities in terms of the nature of content across conversations is expected among the more casual topics. The cumulative density function of *uh* frequency therefore roughly aligns with this topic split.

The between-speaker correlations are reported in Table 1. Here the correlations are calculated from a subset of the analysis sample in which both sides of the conversation appeared in the pool of speakers. Three conditions are separated to account for the potential gender effect on accommodation. This results in a subsample consisting of 685 male-male conversations, 885 female-female conversations and 675 male-female conversations.

Table 1. Correlation of filled pause frequency in conversation between speakers controlled for gender (\*:  $p < 0.05$ ).

	<i>um</i>	<i>uh</i>
Male-male conversation	0.113*	0.368*
Female-female conversation	0.103*	0.205*
Male-female conversation	0.056	0.192*

In all conditions, higher correlations are found in the frequency of *uh*. The smallest difference is in the case of female-female conversations, where the magnitude is still two times higher. Although the absolute between-speaker correlations are at most moderate, the clear difference between the two fillers nevertheless suggests that *uh* is more likely to be accommodated by the interlocutor, especially when both sides are males. This is consistent with earlier observations that *uh* is subject to more variation compared to *um*.

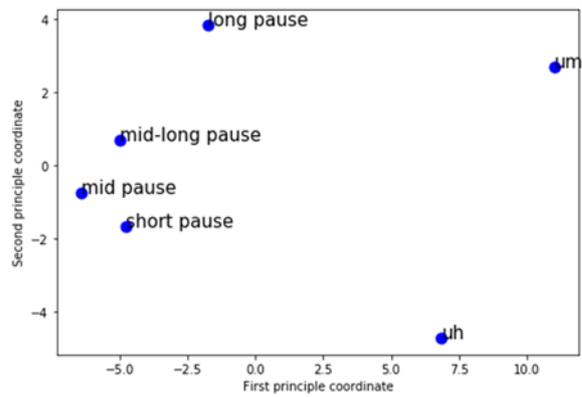


Figure 6. LSA analysis of silent and filled pauses projected to the 2D space using MDS.

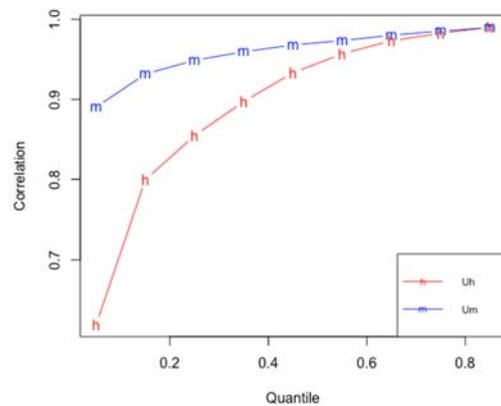


Figure 5. Quantile plot of between topic correlations of the frequency of *um* and *uh*.

The observations presented so far all suggest that *uh* has greater variability in its distribution in spontaneous conversations compared to *um*. One potential explanation for these observations is the functional or meaning difference between the two fillers. This difference may also relate to the syntactic patterns of different speaker groups.

## A potential difference in meaning

In this section, the possibility that the observed difference in frequency variability between the two fillers entails a difference in their meanings is examined through Latent Semantic Analysis (LSA). Essentially, LSA uses the contexts in which a word occurs to represent the ‘meaning’ of the word of interest. Implementations stemmed from this idea have found wide applications in modern NLP applications.

A word2vec model (Mikolov et al., 2013) was trained with a window size of  $\pm 5$  words. Metrical Multidimensional Scaling (MDS) was used to project the clusters in word vector space onto the 2D plane for visualization, as plotted in Figure 6. Silent pauses of varying duration were also coded as

different words to explore the correspondence between filler forms and types of delay. Duration thresholds of 400 ms, 600 ms, and 800 ms were arbitrarily set to distinguish between shorter and longer pauses among all the pauses longer than 150 ms.

The first major difference is between silent and filled pauses along the first principle coordinate. In addition, clear difference between the two fillers can also be found along the second principle coordinate, suggesting a distinction of the contexts in which the two fillers potentially occur. More interestingly, the continuum from short to long silent pauses are dispersed along the second principle coordinate, which appears to parallel with the distinction between *uh* and *um* along the same dimension. Therefore it can be argued that there are some contextual similarities between shorter pauses and *uh*, and between longer pauses and *um*.

## Discussion and conclusion

In this study I presented evidence for an alternative explanation to the previous observation that the choice of *um* and *uh* as hesitation markers displays consistent age-related variations. Through looking at the absolute frequency variation as a function of age, as well as the effect of topic and interlocutor on individual variation, it can be hypothesized that the so-called change in progress can be at least partially explained by a potential meaning or function difference. This difference may be interpreted as reflecting different aspects in the planning process, which correspond to age-related change in utterance structuring and cognition. This potential difference in meaning or function has found support from LSA. However, substantive proposals for how the meaning and function of the fillers are different is not yet able to be proposed. One obvious possibility is to look at variations in other forms of filler words (such as *be like*) in relation with the two forms of filled pause.

## References

- Acton, E. K. 2011. On gender differences in the distribution of *um* and *uh*. *University of Pennsylvania Working Papers in Linguistics* 17(2): 2.
- Cieri, C., D. Miller & K. Walker. 2004. The Fisher Corpus: A resource for the next generations of speech-to-text. In: M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa & R. Silva (eds.), *Proceedings of Language Resources and Evaluation*, Lisbon, Portugal, 69–71.
- Clark, H. H. & J. E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1): 73–111. [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- Corley, M. & O. W. Stewart. 2008. Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass* 2(4): 589–602. <https://doi.org/10.1111/j.1749-818X.2008.00068.x>
- Fruehwald, J. 2016. Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics* 22(2): 41–49.
- Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in spontaneous speech*. London & New York: Academic Press.
- Labov, W. 1994. *Principles of Language Change. Volume 1: Internal Factors*. Oxford: Wiley-Blackwell.
- Levelt, W. J. 1983. Monitoring and self-repair in speech. *Cognition* 14(1): 41–104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Mikolov, T., K. Chen, G. Corrado & J. Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- O'Connell, D. & S. Kowal. 2005. *Uh* and *um* revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research* 34(6): 555–576. <https://doi.org/10.1007/s10936-005-9164-3>
- Rochester, S. R. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research* 2(1): 51–81. <https://doi.org/10.1007/BF01067111>
- Tottie, G. 2011. *Uh* and *um* as sociolinguistic markers in British English. *International Journal of Corpus Linguistics* 16(2): 173–197. <https://doi.org/10.1075/ijcl.16.2.02tot>
- 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. <https://doi.org/10.1163/22105832-00602001>
- Yuan, J. & M. Liberman. 2008. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America* 123(5): 3878. <https://doi.org/10.1121/1.2935783>