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: The structural signaling effect of silent and filled pauses

Author(s): Ralph L. Rose

Abstract: Filled pauses (uh, um) have been shown in a number of studies to have a facilitative effect for listeners, such as helping them better perceive the syntactic structure of ongoing speech. This may be because the extra time afforded by the filled pause gives listeners more time to process the input. Theoretically, then, silent pauses should show a comparable effect. The present study tests this prediction using a grammaticality judgment task following a study by Bailey and Ferreira (2003). Results show that filled and silent pauses have a comparable influence on listeners' grammaticality judgments but further suggest that listeners deem silent pauses as more important and influential markers.

DOI: <https://doi.org/10.21862/diss-09-006-rose>

Citation (JIPA): Rose, Ralph L. 2019. The structural signaling effect of silent and filled pauses. 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, 19–22.

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).

The structural signaling effect of silent and filled pauses

Ralph L. Rose

Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Abstract

Filled pauses (uh, um) have been shown in a number of studies to have a facilitative effect for listeners, such as helping them better perceive the syntactic structure of ongoing speech. This may be because the extra time afforded by the filled pause gives listeners more time to process the input. Theoretically, then, silent pauses should show a comparable effect. The present study tests this prediction using a grammaticality judgment task following a study by Bailey and Ferreira (2003). Results show that filled and silent pauses have a comparable influence on listeners' grammaticality judgments but further suggest that listeners deem silent pauses as more important and influential markers.

Introduction

The signaling capability of filled pauses in English (i.e. *uh/um*) to listeners has been observed in a wide variety of research. Arnold et al. (2003) observed that filled pauses influence listeners' judgments of whether a following noun phrase is a given or new entity in the ongoing discourse. Corley, MacGregor & Donaldson (2006) observed that listeners judged nouns that are immediately preceded by a filled pause as being lower-frequency nouns than those that are not preceded by a filled pause. Brennan and Schober (1999) observed that listeners recover faster when a speech repair is accompanied by a filled pause than when it is not and Bailey and Ferreira (2003) observed that filled pauses influenced listeners' interpretation of the structure of a sentence.

One account of these observations (cf. Bailey & Ferreira, 2003; Corley, MacGregor & Donaldson, 2006; Corley & Stewart, 2008) is what will be referred to here as the *extra time hypothesis*: The additional time afforded to listeners while a filled pause is occurring—as opposed to the contrary case when no filled pause occurs—allows additional linguistic processing to occur, facilitating the listeners' accurate processing of the linguistic input. Evidence consistent with this hypothesis has been obtained in several of the above studies using other sounds in place of filled pauses such as bells, door knocks, or car horns.

However, if the extra time hypothesis is the correct explanation for the various observations, then one would expect that silent pauses—which also afford additional time to the listener for processing—

placed where filled pauses might have occurred should show the same effect.

This paper reports on a test of this prediction. The following section reviews previous research on silent and filled pauses and describes the signaling and extra time hypotheses in more detail. The experimental section describes the grammaticality judgment and reaction-time experiment used to evaluate the hypotheses. And the final section discusses the findings and their implications.

Background

Filled pauses

Filled pauses are non-lexical vocalizations uttered by speakers which, by their occurrence, delay the transmission of the linguistic speech signal. In English, where a large proportion of filled pause research has taken place, two basic forms are nearly exclusive: *uh* [ə:] and *um* [ə:m] (MacLay & Osgood, 1959; Goldman-Eisler, 1961; Mahl, 1987; Shriberg 1994; Clark & Fox Tree, 2002). Many languages have nearly equivalent forms using the same vowel [ə], if available, or a nearby vowel as in French: *eu*h [œ:], *eum* [œ:m] (Vasilescu, Nemoto & Adda-Decker, 2007). Other languages may have somewhat different forms with possibly more than one syllable as in Japanese: *e-* [ɛ:], *e-to* [ɛ:to] (Maekawa, 2003).

Research on the production and perception of filled pauses in speech has been extensive, covering various academic fields, languages, and investigative paradigms (see Clark & Fox Tree, 2002 for a slightly dated overview). Some of the perceptual research has already been introduced above. Bailey and Ferreira's (2003) work is often cited and underlies the present study. In their work (specifically, their third experiment), they examined the signaling effect of filled pauses relative to clause boundaries in ongoing speech. In a grammaticality judgment task, native English listeners heard sentences as in (1) to (4), in which clause boundaries are marked by brackets.

- (1) [Sandra bumped into the uh uh busboy] and [the waiter got angry].
- (2) [Sandra bumped into the busboy] and [the uh uh waiter got angry].
- (3) [While [the man hunted] the uh uh deer ran into the woods].
- (4) [While [the man hunted] the deer uh uh ran into the woods.]

When the filled pause was at a clause boundary (actually, just after the boundary) as in (2) and (3), listeners were more likely to judge the whole sentence as grammatical than in cases with sentences with a filled pause at a non-boundary as in (1) and (4). Bailey and Ferreira interpreted this as a signaling effect: Listeners interpreted the filled pause as a signal of greater cognitive effort by the speaker, such as that which might be caused by planning a larger speech constituent such as a clause. Thus, they described the filled pauses in (2) and (3) as “good” signals—consistent with the syntactic structure—and those in (1) and (4) as “bad” signals—inconsistent with the structure.

Comparable results were obtained by Bailey and Ferreira when using environmental noises in place of filled pauses. Other researchers have also observed similar effects (e.g. Corley, MacGregor & Donaldson, 2006). Thus, a broader explanation for the observations is the extra time hypothesis. Under this hypothesis, it is the extra time that is afforded by the presence of the filled pause that allows the listener to process the ongoing speech. Hence, when the filled pause is in a position that is consistent with larger constituent boundaries, the listener has time to correctly parse the sentence: the extra time is facilitative. Under this hypothesis, the filled pause is not necessarily a signal, but rather merely an irrelevant filler of time. It is that extra time during which listener processing occurs that causes the observed effects.

Silent pauses

Silent pauses are silent periods which are of unusual duration during a speaker’s on-going speech. This excludes short silent periods which may be attributable to breathing or articulatory gestures. Because it is somewhat difficult to determine what counts as “unusual”, a common approach in speech studies is to use a somewhat arbitrary threshold. However, even this threshold value has been set widely, as short as 50 ms to as long as 1 sec (De Jong & Bosker, 2013). In recent years, a threshold value around 250–300 ms is common.

Silent pauses, if counted as discrete units of speech, are among the most common tokens in speech corpora (e.g. 3.8 per 100 words in the LOCNEC corpus; Gilquin, 2008). Furthermore, they may play a crucial role in speech perception. Reich (1980) observed that listeners recalled propositions more accurately when silent pauses were at clause boundary rather than non-boundary positions. These findings might also be explained by the extra time hypothesis: The extra time afforded by the silent pause allows the listener to properly process the structure of the ongoing speech.

Thus, one might predict that filled and silent pauses should show equivalent effects. Yet, another view might be that perhaps the filled pause (or other overt acoustic evidence such as an environmental sound) is not a signal, per se, but simply heightens attention (as discussed in Corley & Stewart, 2008; MacGregor, 2008), perhaps even to the presence of the concurrent delay. This could cause the listener to capitalize on the afforded extra time more effectively.

The present study seeks to investigate these possibilities by replicating Bailey and Ferreira’s (2003) study but extending it by comparing the influence of both filled and silent pauses.

Experiment

Materials

The present study used the original stimuli from Bailey and Ferreira (2003) as shown in (1) to (4) with some important changes. In their study, they placed the filled pause between a definite article and its head noun: *the uh uh waiter*; hence, one token away from the relevant clause boundary. This was because their work was building on earlier work on the head noun effect (Ferreira & Henderson 1991) and thus compared these cases to such cases as *the short and pudgy waiter* or *the waiter who was short and pudgy*. However, following their own argumentation, the signaling effect should occur if the filled pause is right at the clause boundary. Thus, in present study, the placement of the filled pauses was as in (5) to (8). Furthermore, only a single filled pause was used rather than the double used in their study. Silent pause stimuli were made by simply replacing the filled pauses with silence.

- (5) [Sandra bumped into uh the busboy] and [the waiter got angry].
- (6) [Sandra bumped into the busboy] and [uh the waiter got angry].
- (7) [While [the man hunted] uh the deer ran into the woods].
- (8) [While [the man hunted] the deer uh ran into the woods.]

Their original set of stimuli including coordination (i.e. (5)–(6)) and subordination (i.e. (7)–(8)) constructions was extended to make a total of 90 stimuli. 100 filler stimuli were also used, half of which were ungrammatical in some way.

The stimuli were recorded with a native English speaker’s voice (the author’s). The filled and silent pauses were acoustically manipulated to ensure that both were a consistent 500 ms long in all stimuli.

Procedure

After completing a consent form, participants were seated in a quiet room with a computer and a pair of comfortable headphones. Stimuli were presented in a randomized manner for each participant using Superlab 4.0 by Cedrus. During each audible stimulus, participants saw a cross “+” fixation symbol. Afterward, they were prompted to judge whether the sentence they heard was grammatical or ungrammatical by pressing a button on the keyboard. Participants were given an explanation of how to judge grammaticality and explicitly told not to judge sentences as ungrammatical merely because of any disfluencies. They also were given sixteen practice items with reinforcing feedback after each. The entire experiment took most participants about 25 minutes.

The recorded data included participants’ grammaticality judgments as well as their reaction times, being the time from the offset of the stimulus sentence to the onset of their key press. The data were analyzed using mixed effects modeling with the nlme package (v. 3.1-128) in R (v. 3.3.2). Fixed effects were signal status (consistent, inconsistent; corresponding to Bailey & Ferreira’s, 2003 “good” and “bad” conditions, respectively), gap type (silent, filled) and construction (coordination, subordination). Participants were added to the model as a random effect.

Results

30 university students who were native speakers of North American English participated in the experiment and received US\$10 in remuneration.

As illustrated in Figure 1, participants judged the sentence as grammatical more often when the pause was in a consistent signal position (i.e. (6) and (7) above) than when in an inconsistent position (i.e. (5) and (8)) [$t(206) = 8.5, p < 0.001$]. Furthermore, an interaction between signal status and gap type [$t(206) = 3.2, p < 0.005$] reveals that in the inconsistent signal condition, participants judged the silent pause case less grammatical than the filled pause case. Overall, the mixed effects model explains 29.3% of the variance (marginal R^2).

A significant difference in construction reveals that there is a difference between coordination and subordination stimuli. Hence, Figure 2 shows the breakdown by construction. While the basic pattern of results is unchanged, it is clear that the results are more pronounced with subordination stimuli. Furthermore, post-hoc comparisons of the subordination stimuli show that there is even a difference between the pauses in the consistent

condition [$t(29) = 2.1, p < 0.05$]: Participants judged the filled pause stimuli as grammatical less often than the silent pause stimuli. This is a reversal of the pattern seen in the inconsistent signal case.

Figure 1. Proportion of stimuli judged grammatical (normed against individuals’ judgments). Error bars represent 95% confidence intervals.

Figure 2. Proportion of stimuli judged grammatical (normed against individuals’ judgments) for coordination and subordination stimuli. Error bars represent 95% confidence intervals.

As for the reaction time data (see Figure 3), the trends are in a direction that parallels the grammaticality judgment data: The stimuli in the inconsistent signal condition elicit longer reaction times than those in the consistent condition. Furthermore, the silent pause stimuli yield a slightly longer reaction time than the filled pause stimuli. However, none of these trends reach significance and the marginal R^2 is only 1.2%. The only difference is between the coordination and subordination stimuli with the subordination stimuli showing longer reaction times overall [$t(2127) = 4.7, p < 0.001$]. But this difference is not particularly pertinent to the present study.

Figure 3. Mean grammatical judgment latency as reaction time (normed against individuals' reaction times). Error bars represent 95% confidence intervals.

Discussion

Results replicate Bailey and Ferreira's (2003) study with respect to filled pause stimuli: Participants judge them less grammatical when in an inconsistent signal position. But the results here further show that a silent pause in an inconsistent signal position is worse, and also that a silent pause in a consistent position is better. The extra time hypothesis is not sufficient to explain this. A partial explanation might come from the idea that the acoustics of the filled pause (like environmental noises) raises the listeners' attention even during the delay time. But this would not quite explain why the silent pause in the consistent condition of subordination stimuli actually yields the highest rate of grammaticality judgment.

In short, the results here suggest that listeners are processing silent pauses and filled pauses each as discrete units of communication and that each signals something a little different. Or, perhaps that they signal the same thing, but at different levels of intensity. But even if so, it would appear that silent pauses—although acoustically nil—are more influential signals—even louder signals—than filled pauses.

Acknowledgments

This work has been partially funded by a Japan Society for the Promotion of Sciences (JSPS) Grant-in-aid (Scientific Research (C), Project #15K02765). I am grateful to Dr. Bonita Miller and Spring Arbor University for their help in carrying out the experiment and also to numerous individuals who have given helpful advice.

References

- Arnold, J., M. Fagnano & M. K. Tanenhaus. 2003. Disfluencies Signal Thee, Um, New Information. *Journal of Psycholinguistic Research* 32(1): 25–36. <https://doi.org/10.1023/A:1021980931292>
- Bailey, K. G. D. & F. Ferreira. 2003. Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language* 49(2): 183–200. [https://doi.org/10.1016/S0749-596X\(03\)00027-5](https://doi.org/10.1016/S0749-596X(03)00027-5)
- Brennan, S. E. & M. F. Schober. 1999. Uhs and interrupted words: The information available to listeners. In: *Proceedings of Disfluency in Spontaneous Speech*, 30 July 1999, Berkeley, CA, 19–22.
- Corley, M., L. J. MacGregor & D. Donaldson. 2006. It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition* 105(3): 658–698. <https://doi.org/10.1016/j.cognition.2006.10.010>
- 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>
- Clark, H. & J. 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)
- De Jong, N. H. & H. R. Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In: R. Eklund (ed.), *The 6th Workshop on Disfluency in Spontaneous Speech*, 21–23 August 2013, Stockholm, Sweden, 17–20.
- Ferreira, F. & J. M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30(6): 725–745. [https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
- Gilquin, G. 2008. Hesitation markers among EFL learners: Pragmatic deficiency or difference? In: J. Romero-Trillo (ed.), *Pragmatics and Corpus Linguistics: A Mutualistic Entente*, 119–150. Berlin: Mouton de Gruyter.
- Goldman-Eisler, F. 1961. A Comparative Study of Two Hesitation Phenomena. *Language and Speech* 4(1): 18–26. <https://doi.org/10.1177/002383096100400102>
- MacGregor, L. J. 2008. *Disfluencies affect language comprehension: evidence from event-related potentials and recognition memory*. Ph.D. dissertation, The University of Edinburgh.
- Maclay, H. & C. Osgood. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15(1): 19–44. <https://doi.org/10.1080/00437956.1959.11659682>
- Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its Design and Evaluation. *Proceedings of SSPR*, 13–16 April 2003, Tokyo, Japan, 7–12.
- Mahl, G. 1987. *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Reich, S. 1980. Significance of Pauses for Speech Perception. *Journal of Psycholinguistic Research* 9(4): 379–389. <https://doi.org/10.1007/BF01067450>
- Shriberg, E. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. dissertation, University of California at Berkeley, USA.
- Vasilescu, I., R. Nemoto & M. Adda-Decker. 2007. Vocalic Hesitations vs Vocalic Systems: A Cross-Language Comparison. In: J. Trouvain (ed.), *Proceedings of 16th International Congress of Phonetic Sciences*, 6–10 August 2007, Saarbrücken, Germany, 1101–1104.