A CROSSLINGUISTIC CORPUS STUDY OF SILENT AND FILLED PAUSES: WHEN DO SPEAKERS USE FILLED PAUSES TO FILL PAUSES?

Ralph Rose[1] and Michiko Watanabe[2]

[1]Waseda University, [2]National Institute for Japanese Language and Linguistics
[1]rose@waseda.jp, [2]watanabem@ninjal.ac.jp

ABSTRACT

During spontaneous speech production, speakers inevitably pause silently due to various causes. Often, speech resumes normally afterward without message interruption. But sometimes, silence is terminated with a filled pause (English *uh*/*um*, Japanese *e:*/*e:to*). The present study examines whether such filled pauses may be regarded as "pause fillers": that is, if a speaker pauses longer than some (e.g., culturally-determined) silence threshold, they are likely to fill the silence with a marker (here, a filled pause). Four questions are examined: (1) do speakers use filled pauses as pause fillers, (2) does this differ between native English and Japanese, (3) does Japanese speakers' silence threshold differ between their first and second languages, and (4) are such differences modulated by second language proficiency? Results from a study of two monologic speech corpora suggest that the answer to all questions is "yes", but only with utterance-internal rather than -boundary pauses.

Keywords: silent pauses, filled pauses, speech prosody, speech corpora, hesitation phenomena

1. INTRODUCTION

Speakers engaging in unscripted speech will inevitably find themselves from time to time unable to continue their speech immediately. This could happen after the end of one utterance or even mid-utterance. In either case, they may become temporarily silent at this point before resuming. Sometimes, this silence may be terminated with the beginning of a new utterance, and other times, the termination will be with a resumption of the current utterance. Sometimes, the termination will occur with a filled pause of a conventional form in the language of their speech (e.g., *uh*/*um* in English) before speech resumes normally. This filled pause may be seen as "filling" some (further) silence that would have otherwise occurred in its place. Furthermore, the motivation for its use could be that the speaker realizes their silence will exceed some threshold for normal silence. This hypothesis—herein called the "pause filler hypothesis" and motivated in part by work as far back as Maclay and Osgood [8] inter alia—is partially examined in the present work.

Under this hypothesis, the use of filled pauses as pause fillers could be modulated by language. That is, for speakers of languages where cultural norms allow lengthy silence, the silence threshold could be longer and the onset of pause fillers might be delayed compared to speech in other languages whose cultures evaluate silence more negatively. Alternatively, nonnative speakers of a language with a different cultural value of silence than that of their native language may show some variation in their speech production as a transfer effect. The present study examines pause fillers within these crosslinguistic comparisons using two speech corpora.

2. BACKGROUND

2.1. Filled pauses

Filled pauses are conventionalized phonemic forms that carry no propositional value (though they may have a pragmatic purpose) and delay message transfer. In English, typical forms are *uh* [əː] and *um* [əm] [3, 6, 8, 10, 16]. In Japanese, forms are more varied, but by far the most common are *e-* [ɛː] and *e-to* [ɛːto] [9].

The label "filled pause" is somewhat controversial, with one point of contention being their nature. Some researchers (cf., [3, 17]) argue that they are not "pauses" at all but are intentionally-chosen devices used to communicate expectations about problems in language production. Some alternative names for the phenomenon thus include "hesitation fillers" and "ah-phenomena". Other researchers argue against the notion of filled pauses as words or intentionally selected devices (cf., [4, 5]), instead regarding them as something more like nonverbal tics. The present study makes use of the terms "fill" and "pause" as a terminological convenience, but is not committed to any particular side in the above debate. The somewhat intermediate view taken herein views filled pauses as occurring when a speaker realizes the silence threshold will be exceeded and then uses them to fill (or replace) what would have been further silence.

2.2. Silence threshold

Numerous factors are hypothesized to influence the silence threshold. Previous research shows that the

occurrence of silent and filled pauses is influenced by the discourse-syntactic context: In general, pauses are longer and more frequent at major boundaries than at minor boundaries (e.g., in English [3, 18] and in Portuguese [13]). It is conceivable that the silence threshold is longer at such major boundaries: Speakers give themselves longer to prepare a major following constituent, and listeners may also cooperatively allow them to do so. But pauses at minor or non-boundary locations may be regarded as unwarranted and hence the allowable silence may be shorter, lest it generate confusion in communication.

But this may vary crosslinguistically. In Japanese, for example, silent pauses show a distribution similar to that described above, but filled pauses do not [20]. This could be culturally driven. Japanese (and some other Asian languages) has been described by sociologists as being more tolerant of silence (cf., [7, 11, 12]). This could influence the Japanese speaker's threshold length and, hence, their use of filled pauses as pause fillers.

Yet another influence on silence threshold could arise in the context of second language (L2) speech. Nonnative speakers may be less proficient in the language and therefore have no choice but to pause longer and perhaps in what might otherwise be unusual locations for the native (L1) speaker. They might therefore show a longer silence threshold than they normally would (cf., transfer effects as described in [15, 19]). On the other hand, as they become more proficient in the second language, they might show a silence threshold that is more in line with that of the target language (e.g., Japanese learners of English might begin to use a shorter threshold, particularly at major boundaries).

3. METHOD

The present work aims to answer four research questions, as follows.

(1) Do speakers use filled pauses as pause fillers?
(2) Does this differ between native languages?
(3) Does speakers' silence threshold differ between their first and second languages?
(4) Are such differences modulated by second language proficiency?

3.1. Spontaneous speech corpora

For the purpose of answering these questions, L1 and L2 speech in two different speech corpora were analyzed. While these two corpora are not directly related to one another, they have comparable elements in the context of the present work.

3.1.1. Corpus of Presentations in English

The Corpus of Presentations in English (COPE, [21]) consists of narrative speech by 20 university-aged native speakers of North American English. Speakers were given ten minutes to make notes and prepare to give an otherwise unscripted ten-minute talk on their most memorable experience. As a whole, the corpus consists of 41,062 words in 3.8 hours of speech. This includes 7,156 silent pauses and 1,442 filled pauses.

3.1.2. Crosslinguistic Corpus of Hesitation Phenomena

The Crosslinguistic Corpus of Hesitation Phenomena (CCHP [14]) is a corpus of speech in response to three different tasks: reading aloud, picture description, and topic narrative. 35 university-aged native speakers of Japanese performed each task for approximately three minutes in *both* Japanese (their L1) and English (their L2). For the present work, only the unscripted speech portions (excluding reading aloud) were used. This consists of 40,870 words in 9.2 hours of speech and includes 13,648 silent pauses and 3,680 filled pauses. The corpus also contains a general L2 proficiency estimate (on a course scale, 1-7) based on self-reported performance on standardized tests (typically focused on receptive skills), living abroad experience, and self-assessment of proficiency.

3.2. Extracted data

To test the basic pause filler hypothesis, the relevant comparison is between the duration of standalone silent pauses and the total duration of pauses starting with silence and then terminating in a filled pause. Hence, for all silent pauses in both corpora, duration measurements were taken of silent pauses as well as silent plus filled pauses using Praat [2]. For all of these pauses, also recorded was whether they occurred at an utterance boundary, or utterance-internally. Finally, the CCHP L2 proficiency estimates were used to separate speakers into two broad proficiency groups: *low* and *high* (17 speakers in each group; one speaker for whom no L2 proficiency information was available was excluded from the analysis).

In order to evaluate research questions (1) and (2), the L1 English data from COPE and the L1 Japanese data from CCHP were used. To evaluate research questions (3) and (4), the L1 Japanese data and L2 English data from CCHP were used. Statistical tests were performed using mixed effects modeling using duration, language, and L2 proficiency group as fixed effects and participants (speakers) as random effects. Because the distribution of pause duration is log-normal, the `glmmPQL()` function in the `MASS`

Figure 1: Duration (with 95% confidence intervals) of standalone silent pauses (dark gray) and silent followed by filled pauses (light gray) at utterance boundaries in L1 English (COPE) and L1 Japanese (CCHP) at left and in L1 Japanese (CCHP) and L2 English (CCHP) for low and high L2 proficiency speakers at right.

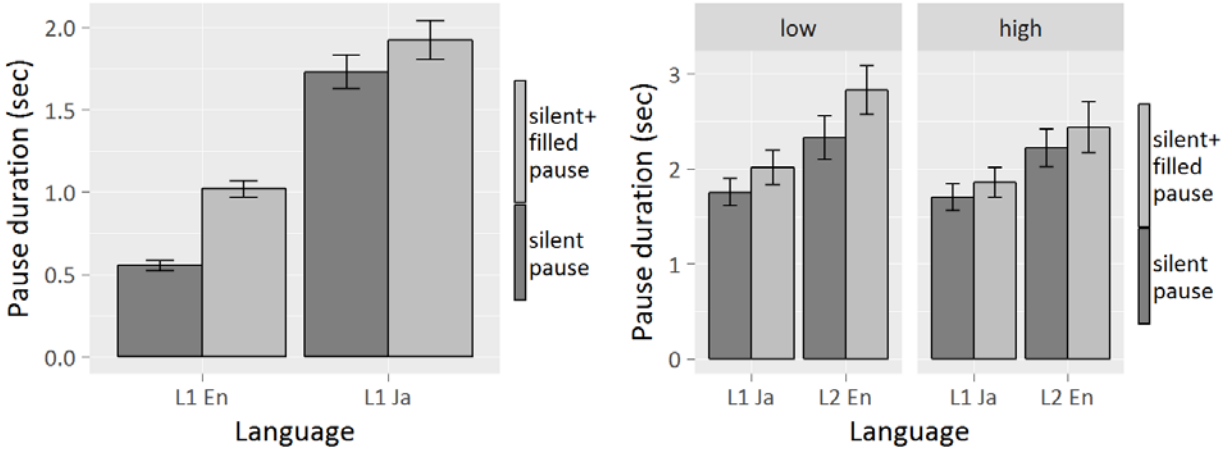
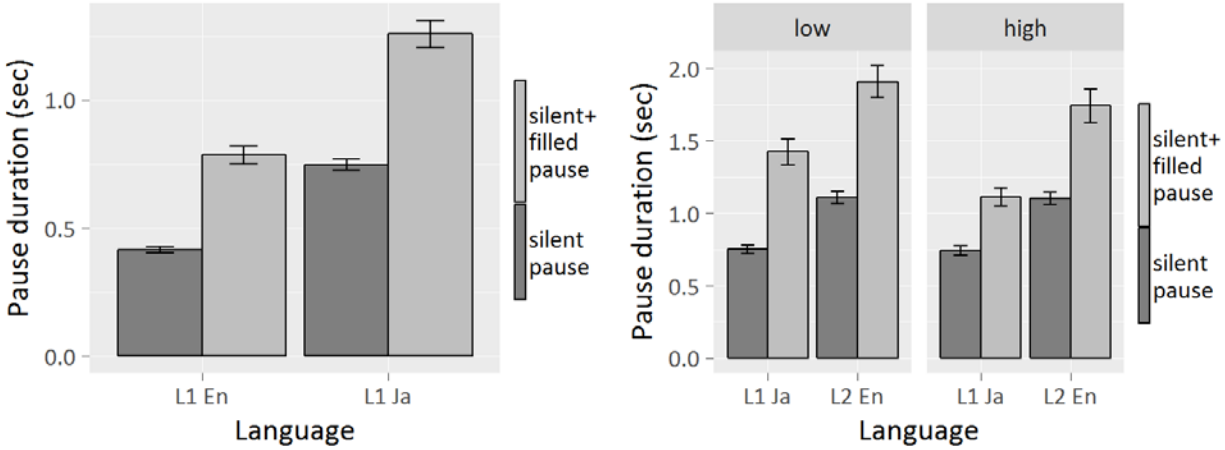


Figure 2: Duration (with 95% confidence intervals) of standalone silent pauses (dark gray) and silent followed by filled pauses (light gray) at utterance-internal locations in L1 English (COPE) and L1 Japanese (CCHP) at left and in L1 Japanese (CCHP) and L2 English (CCHP) for low and high L2 proficiency speakers at right.



package (ver. 7.3-45) in `R` (ver 3.3.2) was used because it provides the option to specify the GLM family (i.e., `gaussian(link="log")`).

One further comment is warranted regarding the analysis. In much work, pause duration is known to vary somewhat with speech rate. As a result, many researchers normalize pause duration measurements with respect to speech rate. In the analysis presented here, this was not done: pause durations are the raw durations. This is because the pause filler hypothesis is predicated on the idea that the silence threshold is a shared property among a community of language speakers. Hence, I assume that speakers will observe a common silence threshold duration, in real-time. [Note: Although there is no space to report it here in detail, an alternative statistical analysis based on normalized durations showed the same overall trends that are reported below.]

4. RESULTS

As noted above, different factors may influence the silence threshold at utterance boundary vs. internal locations. Therefore, the results for each of these are presented separately below.

4.1. Utterance boundary pauses

Results (Fig. 1 left) show that utterance boundary pauses are longer in L1 Japanese than in L1 English [$t(52)=19.8$, $p<0.001$] and further that in both languages, silent plus filled pauses are longer [$t(2642)=8.8$, $p>0.001$]. However, a significant interaction between duration and language [$t(2642)=3.8$, $p<0.001$] suggests that the difference between the two pause types is weaker in L1 Japanese. Under this model, the marginal $R^2 = 26.7\%$.

Results (Fig. 1 right) also show that utterance boundary pauses are longer in L2 English than in L1 Japanese [$t(1840)=6.0$, $p<0.001$] and further that in

both languages, silent plus filled pauses are longer [t(1840)=3.7, p>0.001]. However, no other main effect nor interaction was observed and the marginal R^2 = 1.9%.

4.2. Utterance-internal pauses

Utterance-internal pauses show a much clearer pattern of results than do utterance boundary pauses. The internal pauses (see Fig. 2 left) are longer in L1 Japanese than in L1 English [t(52)=8.2, p<0.001] and in both languages, silent plus filled pauses are longer than plain silent pauses [t(10,266)=18.6, p<0.001]. The lack of an interaction between duration and language [t(10,266)=1.2, n.s.] further shows that the difference between the two pause types is consistent across languages. For the model as a whole, R^2 = 29.7%.

Results (Fig. 2 right) show that utterance-internal pauses are longer in L2 English than in L1 Japanese [t(10,353)=15.6, p<0.001] and also that silent plus filled pauses are longer than standalone silent pauses [t(10,353)=18.0, p<0.001]. Furthermore, an interaction between L2 proficiency group and pause type [t(10,353)=3.0, p<0.005] shows that the difference between the pause types is larger for low proficiency than high proficiency speakers. For the model, marginal R^2 = 10.4%

5. DISCUSSION

The present work has sought to examine the pause-filler hypothesis—the idea that speakers who anticipate their ongoing silence to exceed some threshold will fill the pause with a filled pause—with respect to the four questions listed at the start of Section 3 above.

Results from the two crosslinguistic speech corpora suggest that the answer to all four questions is yes, but with some qualifications. First, the difference between the pause types for utterance boundary pauses is weaker in L1 Japanese than in L1 English. Further, the difference is consistent between L1 Japanese and L2 English suggesting that they are weak in both cases, regardless of L2 proficiency. Hence, for utterance boundary pauses, L1 English speakers use filled pauses as pause fillers but L1 Japanese and L2 English speakers do not; or do so only weakly and with a longer silence threshold.

For utterance-internal pauses, however, the difference between pause types is consistently strong across all three language variants, and, it seems, especially strong for low proficiency L2 English speakers. Hence, for utterance-internal pauses, all speakers use filled pauses as pause fillers with silence thresholds showing the following order: L1 English < L1 Japanese < low proficiency L2 English < high proficiency L2 English. Interestingly, the results even show that the low proficiency L2 English speakers actually maintain the longer silence threshold even when speaking in L1 Japanese. This suggests that it could be an aspect of their L1 behavior that determines the nature of their L2 performance, hence, a transfer effect from their L1 [15, 19].

The difference between the utterance boundary and non-boundary cases is interesting in light of the pause filler hypothesis. The weaker effect of pause-filling at utterance boundaries suggests that the motivation to observe the threshold is not as important at when the speaker is preparing a subsequent utterance. This is not inconsistent with previous findings which show that speakers use more and longer filled pauses at major than minor boundaries. It just means that they are not as concerned about filling silence that may occur at those places. But they are more concerned at non-boundary locations where an overly long silent pause might seem odd, or even serve as an incorrect signal that the following constituent might be a major constituent (cf., [1]).

One potential empirical criticism of the present work is the way that the pause-filler notion has been operationalized. Here, the motivator of pause-filling has been the total duration of the silent pause plus its immediately following filled pause. But, an alternate concept of the motivator is the duration of the silent portion *alone*. That is, if the silence itself exceeds some threshold, a filled pause will be initiated. While it is not possible to present the statistical details here, this alternative version of the pause-filler hypothesis was tested. Results actually show the same general pattern of results as already presented here, with a stronger distinction between utterance boundary and utterance-internal pauses (i.e., the former showing almost no significant effects at all).

Finally, while the present research examines filled pauses as pause fillers, it should be noted that the results here do not conclusively confirm the pause-filler hypothesis but are merely consistent with the hypothesis. Further work testing the hypothesis in a controlled experimental paradigm designed to measure the production or even perception of filled pauses as pause fillers is necessary.

Nonetheless, the data presented here is somewhat unique in that most studies of silent and filled pauses have looked at the silence after the filled pause (e.g., [3]). While there are some that look at the silence before filled pauses (e.g., [22]), few (if any) have compared silent and filled pauses to standalone silent pauses using crosslinguistic corpora as reported herein. It is hoped that this data provides a new view on the rather older question of whether filled pauses actually fill pauses.

6. REFERENCES

[1] Bailey, K., Ferreira, F. 2003. Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language* 49, 183-200.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341–345.

[3] Clark, H., Fox Tree, J. 2002. Using uh and um in spontaneous speaking. *Cognition* 84, 73–111.

[4] Corley, M., Stewart O. W. 2008. Hesitation Disfluencies in Spontaneous Speech: The Meaning of um. *Language and Linguistics Compass* 2, 589–602.

[5] Finlayson, I. R. 2014. *Testing the roles of disfluency and rate of speech in the coordination of conversation*. Ph.D. Thesis. Queen Margaret University: Edinburgh.

[6] Goldman-Eisler, F. 1961. A Comparative Study of Two Hesitation Phenomena. *Language and Speech* 4, 18–26.

[7] Klopf, D. W. 1991. Japanese communication practices: Recent comparative research. *Communication Quarterly* 39, 130–143.

[8] Maclay, H., Osgood, C. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15, 19–44.

[9] Maekawa, K. (2003) "Corpus of Spontaneous Japanese: Its Design and Evaluation." *Proceedings of SSPR* Tokyo, 7–12.

[10] Mahl, G. 1987. *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[11] Nishimura, S., Nevgi, A., Tella, S. 2008. Communication Style and Cultural Features in High/Low Context Communication Cultures: A Case Study of Finland, Japan and India. *Proc. of a subject-didactic symposium in Helsinki on Feb. 2, part 2* Helsinki, 783–796.

[12] Ohtaki S, Ohtaki T., Fetters, M. D. 2003. Doctor–patient communication: a comparison of the USA and Japan. *Family Practice* 20, 276–282.

[13] Oliveira, M. 2002. The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. *Proc. of the Third International Conference on Advances in Natural Language Processing (PorTAL '02)* Faro, 43–52.

[14] Rose, R. 2013. Crosslinguistic Corpus of Hesitation Phenomena: A Corpus for Investigating First and Second Language Speech Performance. *INTERSPEECH 2013* Lyon, 992–996.

[15] Selinker, L. 1969. Language transfer. *General Linguistics* 9, 67-92.

[16] Shriberg, E. 1994. *Preliminaries to a theory of speech disfluencies*, Ph.D Thesis. University of California, Berkeley: Berkeley.

[17] Smith, V., Clark, H. 1993. On the course of answering questions. *Journal of Memory and Language* 32, 25–38.

[18] Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485–496.

[19] Tajima, K., Port, R. Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25, 1-24.

[20] Watanabe, M., Kashiwagi, Y., Maekawa, K. 2015. The relationship between preceding clause type, subsequent clause length and duration of silent and filled pauses at clause boundaries in Japanese monologues. *The 7th Workshop on Disfluency in Spontaneous Speech (DiSS 2015)* Edinburgh.

[21] Watanabe, M., Toyama, S. 2017.『日本語話し言葉コーパス』と対照可能にデザインされた英語話し言葉コーパスにおけるフィラーの分布の特徴 [Building "The Corpus of Oral Presentations in English (COPE)" for Contrastive Studies of Disfluencies in English and Japanese: With Some Preliminary Analyses of Filled Pause Distributions]. *NINJAL Research Papers 12*, 141–203.

[22] Womack, K., McCoy, W., Alm, C. O., Calvelli, C., Pelz, J. B., Shi, P., Haake, A. 2012. Disfluencies as Extra-Propositional Indicators of Cognitive Processing. *Proc. of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)* Jeju, 1-9.