

PRESENTATION/POSTER

TOWARDS A DEEPER, UH, UNDERSTANDING OF, UM, L2 FLUENCY AND ITS [750 MS SILENCE] CORRELATES

Katie C. Comeaux, Brock University

Ron I. Thomson, Brock University

Previous research indicates that filled pauses (Fox Tree, 2001) and pauses that occur at clause boundaries (Brennan & Schober, 2001) tend to be less deleterious to listener judgments of intelligibility and comprehensibility than pauses produced clause-internally (Kang, 2010). Beyond their impact on ease of processing for listeners, hesitation phenomena may also lead to negative social evaluations of the speakers, particularly if pausing patterns are outside of listeners' linguistic- or culturally bound expectations. The current study aimed to confirm and extend these previous findings through a careful manipulation of natural L2 English speech stimuli, which were then presented to native speakers of English for evaluation. The L2 speech samples used were originally produced by 10 L1 Mandarin and 10 L1 Slavic talkers in the context of an extemporaneous picture description task. Each sample was then carefully manipulated to arrive at five matched versions that were either free of hesitation markers, included hesitation markers at clause boundaries (e.g., um, uh, or silence), or included hesitation markers placed within clauses. Using 9-point Likert-type scales, twenty listeners rated the speech samples for the speakers' fluency, comprehensibility, intelligence (IQ) and socio-economic status (SES). Findings suggest that unfilled pauses located at clause boundaries have a more positive impact on listeners' judgements of fluency, comprehensibility, intelligence, and SES.

INTRODUCTION

Hesitation phenomena (e.g., filled and silent pauses) are commonly observed in speech produced by both Native Speakers (NSs) and Non-native speakers (NNS). While the underlying cause of some hesitations evidences speakers' attempts to buy time during online planning and processing (i.e, cognitive disfluency), pausing also provides interactional cues to interlocutors. Thus, depending on the location and type of pause used, hesitations can have both facilitative and deleterious effects on listeners' comprehension (Fox Tree, 2001; Maclay & Osgood, 1959). Further, although the inclusion of pauses in spoken language may serve the same purpose for both NSs and NNSs, in second language (L2) speech, pauses often reflect real differences in cognitive fluency. L2 learners often have different knowledge of the target language, relative to NSs, and less efficient access to that knowledge. Consequently, listeners often judge NNS's use of pauses more harshly than those of NSs (Riggenbach, 1991), since hesitations may be taken to evidence an impoverished linguistic system. For example, pausing in second language (L2) speech is known to have harmful effects on the perception of a speaker's oral fluency and comprehensibility (Cenoz, 2000). Further, the way speakers utilize, and listeners perceive, pausing may be a culturally mediated act, meaning that certain pause types may be appropriate in one language context, but problematic in another (Cenoz, 2000; Watanabe, Hirose, Den, & Minematsu, 2008).

In this exploratory study, we build upon existing literature to further examine the extent to which (a) judgments of NNS's speech are influenced by the pause type utilized, and (b) whether a pause's location also impacts listener judgements of NNS's fluency, comprehensibility, intelligence and socio-economic status (SES), in a Canadian context.

Background

Despite several decades of research examining pausing phenomena, few studies have attempted to discriminate which pausing features are most detrimental to L2 communication, and of those studies, only a handful have utilized experimental methodologies (e.g., Bosker, Pinget, Quené, Sanders, & De Jong, 2013; Bosker, Quené, Sanders, & De Jong, 2014a; Bosker, Quené, Sanders, & De Jong, 2014b; Derwing, Munro, & Thomson, 2008; Kang, 2010; Kahng, 2018). What is clear is that the location and type of pause encountered in L2 speech matters to listeners.

Previous research reports strong agreement that pauses located at clause boundaries (CB) are less harmful to listener comprehension than pauses located clause internally (CI) (Lennon, 1990; Maclay & Osgood, 1959; Riggenschach, 1991; Watanabe et al. 2008). For example, it may be more difficult for listeners to process the sentence, "A man and woman bumped *uh* into each other and fell down", than the sentence "A man and a woman bumped into each other and *uh* fell down". Differences in the ease of processing such sentences are assumed to result from listeners' perceptions that pauses located at CB are expected, and less disruptive in comparison to pauses produced at CI locations (Maclay & Osgood, 1959; Riggenschach, 1991). Lennon (1990) reports that pauses located at CB's are perceived as being shorter than their CI counterparts of the same length. Additionally, it has been found that pauses located CI can create syntactic ambiguities (Watanabe et al., 2008). Bailey and Ferreira (2003) found that listeners struggle in understanding compound sentences when disfluencies follow head nouns (e.g., "The man *uh* went to the store and bought some booze"). With agreement across previous studies, it appears that pauses located at clause boundaries are less deleterious than pauses located clause internally.

Previous research has also found evidence to suggest that filled pauses differentially affect listeners' perception relative to unfilled pauses, although there is not complete agreement in these results (Blau, 1991; Brennan & Schober, 2001; Brennan & Williams, 1995; Clark & Fox Tree, 2002; Kang, 2010; Riggenschach, 1991; Watanabe et al., 2008). Kang (2010) found that unfilled pauses (UP) negatively influenced listener judgements of comprehensibility and accentedness, whereas filled pauses (FP) had no effect. In addition to listener judgements, FPs were found to lead to faster response times (Watanabe et al., 2008), and faster word recognition when utterances and words were following FPs (Brennan & Williams, 1995). Although there are some studies which suggest that UPs are less detrimental to listener processing than FPs (Cenoz, 2000; Clark & Fox Tree, 2002), the majority of the available evidence suggests that FPs may be less harmful to listener judgements than UPs.

While the present study seeks to further examine how pause location and types affect listener judgements of fluency and comprehensibility, it also examines the impact of pause location and pause type on social judgements of intelligence and SES. Previous research suggests that listeners are prone to making negative social judgements of foreign accented speech (Davila, Bohara, & Saenz, 1993; Munro, 2003; Munro & Derwing, 1995; Rubin, 1992). Davila et al. (1993) found that

listeners perceived accented speakers as having lower incomes. Additionally, Rubin (1992) found that the lectures of university teachers who were perceived as having foreign accents were less understandable to students. Such social evaluations may, at least in part, be triggered by some measurable feature or features of the speech signal. As pausing phenomena vary across languages and cultures, the utilization of certain pausing strategies might then contribute to the perceived accentedness of a speech signal and may further contribute to negative social evaluations of NNSs.

In light of the previous research briefly described above and the questions raised therein, we propose two research questions to guide the current study.

Research questions

1. Do the locations of pauses (clause internal vs. at clause boundary) and the types of pauses (filled vs. unfilled) differentially affect listeners' perceptions of L2 fluency and comprehensibility? If so, which pause locations and types have stronger negative effects?
2. Do L2 English speakers' pause locations and pause types influence listeners' impression of the speakers' intelligence and socio-economic status?

METHODOLOGY

Participants

Twenty English NS listeners were recruited via conspicuously placed posters at a university in Southern Ontario. The listeners were primarily undergraduate students and ranged in age from 19-53 (mean age 25.4), and comprised eight males, eleven females and one non-binary person. Most of the participants were born and had resided in Southern Ontario for most of their lives. Fifteen of the participants reported having studied one or more second languages, however only two reported being fluent in a second language.

Materials

The speech samples utilized in this study were modified from samples used in Isaacs and Thomson (2013), where 20 beginner level, adult English as a second language (ESL) speakers (ten Mandarin L1; ten Russian L1) completed an eight-frame picture description task illustrating a humorous story about a man and a woman who mixed up their suitcases while traveling in a big city (Derwing et al., 2004). Isaacs and Thomson (2013) extracted the first 20 seconds of speech from each learner's description for analysis.

For the purposes of the current study, the 20-second L2 English speech samples used in Isaacs and Thomson (2013) were edited using Wave Pad software to synthesize five pausing conditions: 1) filled pauses at clause boundaries (FPCB); 2) filled pauses clause internally (FPCI); 3) unfilled pauses at clause boundaries (UPCB) 4) unfilled pauses located clause internally (UPCI); and 5) versions that are completely free of filled and unfilled pauses, that is hesitation free (HF). Versions with pauses (filled or unfilled) had between 1 and 3 pauses, depending on the number of clauses available for manipulation. Filled pauses were harvested from those produced by the same speaker within their unmodified speech samples and copied to the desired locations. Unfilled pauses were

also based on each sample speaker's natural unmodified productions, with the same background noise as the rest of the recording, but moved when necessary to create the desired experimental conditions. Silent pauses were defined as any silence longer than 200 ms, a threshold previously identified as disruptive (Goldman-Eisler, 1961). Pause length in the UP conditions were not matched to FP counterparts, in an effort to maintain naturalness. It is possible that pause length could have a deleterious effect on listener judgements, however this requires further investigation in future studies. HF speech samples were modified to remove all filled pauses and all unfilled pauses over 200 ms. In total, these five manipulations resulted in 100 modified speech samples (see Table 1 for examples of each condition).

Table 1

Pausing conditions compared in this study

TYPE	ABBREVIATION	EXAMPLE
Filled pauses clause internal	(FPCI)	“Then they hit uh each other - their suitcase fell down”
Filled pauses at clause boundaries	(FPCB)	“Then they hit each other - uh - their suitcase fell down”
Unfilled pauses clause internal	(UPCI)	“Then they hit <silent pause> each other - their suitcase fell down”
Unfilled pauses at clause boundaries	(UPCB)	“Then they hit each other <silent pause> their suitcase fell down”
Hesitation free versions	(HF)	“Then they hit each other. Their suitcase fell down”

Procedure

Two questionnaires containing 9-point Likert-type scales were utilized to measure listener judgements of fluency, comprehensibility, intelligence, and SES as these scales are known to be reliable measures of listener perception (Derwing, Munro, & Thomson, 2008). The speech measures and the social measures were separated across two questionnaires with comprehensibility and intelligence paired, and fluency and SES paired to avoid the potential of cross-construct influence had we paired speech or social constructs together.

Rating sessions took place in a quiet room in the university. The definitions for each construct were provided prior to the rating sessions to ensure common understanding. Following these brief explanations, two practice items were played and rated to ensure the participants had the opportunity to ask questions and become comfortable with the task. The 100 randomized test items were played twice, once to rate comprehensibility and intelligence and once to rate fluency and

SES. Raters were given a five-minute break between rating sessions to reduce fatigue. The total time to conduct the experiment was one hour and twenty minutes.

Listener judgments were analyzed for interrater reliability using Chronbach's Alpha. Additionally, descriptive statistics and ANOVAs were conducted to identify any differences across pausing conditions, for fluency, comprehensibility, intelligence and SES. Correlations between pausing conditions for each of these constructs were also conducted.

RESULTS

Chronbach's alpha coefficients indicated strong interrater agreement on ratings for each construct: Fluency, .900; Comprehensibility, .952; Intelligence, .937; and Socio-economic status, .942. Thus, raters' scores were collapsed to examine mean ratings for each sample.

A repeated measures ANOVA provided a mixed result. Pauses appear to most impact Comprehensibility ratings, with significant effects on the ratings for both pause location [$F(1, 19) = 5.379, p = 0.032, \eta^2 = 0.221$], and pause type [$F(2, 38) = 6.138, p = 0.005, \eta^2 = 0.244$]. Raters preferred samples with pauses at clause boundaries, and preferred UPs over FPs. For Fluency, only pause type [$F(2, 38) = 17.402, p < 0.000, \eta^2 = 0.478$] was found to have a significant effect on judgments. That is samples with FPs were perceived as significantly less fluent than samples with UPs. Similarly, for both IQ [$F(2, 38) = 5.819, p = 0.006, \eta^2 = 0.234$] and SES [$F(2, 38) = 4.526, p = 0.017, \eta^2 = 0.192$], again, pause types matter (raters prefer UPs) but not pause location. It is also worth noting that the ANOVA revealed no significant interactions between pause type and pause location within any construct. Effect sizes for pause type ranged from small to medium across constructs.

Descriptive statistics for the means (see Figure 1) illustrate these general patterns for all speech constructs (i.e., fluency, comprehensibility, IQ and SES). Hesitation free samples (i.e., those that had no filled nor silent pauses longer than 200 ms) were most preferred. Samples containing filled pauses were rated more negatively than samples containing unfilled pauses. Within the filled pause conditions, pauses located clause internally had either no impact or a more negative impact on ratings than pauses located at pause boundaries. Given the relatively small sample sizes, we do not feel a greater examination of variance patterns is warranted for this exploratory study.

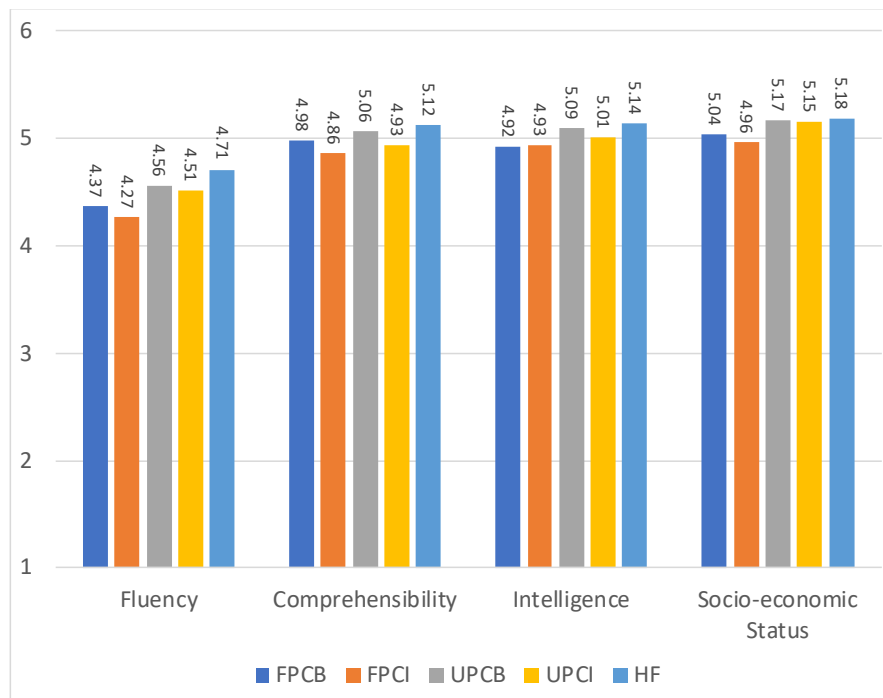


Figure 1. Rater means across pausing conditions (Filled pauses at clause boundaries (FPCB); Filled pauses clause internally (FPCI); Unfilled Pauses at clause boundaries (UPCB); Unfilled pauses clause internally (UPCI); and Hesitation free (HF) speech.

We also examined the extent to which individual raters, on average, preferred speech samples with UP over FP. In 70% of cases the unfilled pause version of a sample was preferred (i.e., rated more favorably). In 25% of cases the filled pause version was preferred. And in 5% of the cases there was no preference. This demonstrates that raters were largely consistent with each other and were affected by the same speech features in similar ways.

Across all constructs and pause types, there was a difference in ratings across speakers. For instance, the Mandarin speakers consistently received lower ratings than the Russian speakers in every construct and pausing condition. This is likely not meaningful but reflects the fact that the Mandarin speakers, as a group, were less proficient in English and therefore less fluent. While it might also be possible that the nature of the filled pauses used by the Mandarin vs. the Slavic speakers impacts ratings, and this should not be ruled out, the speech rate of the Mandarin speakers was slower (see Isaacs and Thomson, 2013), and so the effect of L1 cannot be disentangled from their overall temporal fluency.

Pearson correlation coefficients were also conducted between the speech measures (mean fluency and comprehensibility) and the social judgments (mean IQ and SES) (See Table 2). Correlations between fluency and the social dimensions were extremely high in both filled pause and unfilled pause contexts. Similarly, correlations between comprehensibility and the social dimensions were extremely high. It is worth recalling that fluency and IQ were not rated simultaneously, nor were comprehensibility and SES. Therefore, the strength of these correlations cannot be attributed to the influence of rating these constructs at the same time.

Table 2

Pearson correlations between mean speech measures and social measures

	IQ Filled pause condition	SES Filled pause condition
Fluency Filled pause condition	.952**	.963**
Comprehensibility Filled pause condition	.958**	.925**
	IQ Unfilled pause condition	SES Unfilled pause condition
Fluency Unfilled pause condition	.992**	.973**
Comprehensibility Unfilled pause condition	.965**	.913**

DISCUSSION

The current study found that pause type significantly impacts listener judgements of fluency, comprehensibility, intelligence and SES of L2 speakers. Filled pauses, regardless of their location, had a significant negative effect on judgements of fluency and comprehensibility. This is an interesting finding, as some previous research suggests that unfilled pauses may be more deleterious to the listener. One possible explanation for this finding could be the length of pauses, which were, in an effort to maintain naturalness of speech samples, not strictly controlled for in this study. Had pause durations for FP and UP conditions been identical, a different result may have been obtained. For example, a long silent pause broken up by a filled pause may be preferable to a long silent pause without a filler breaking it up. Since we shortened the length of overly long silent pauses after removing fillers, we cannot answer this question. Future research should consider investigating the duration threshold at which silent pauses may become more problematic for listeners than the same amount of silence interrupted by filled pauses.

The location of pauses seemed to matter less than pause type, with a significant negative effect for comprehensibility only, when either filled or unfilled pauses were heard within clauses. The lack of significance for pause location, outside of the comprehensibility construct, may be due to the fact that in order to judge for comprehensibility, listeners need to attend to the meaning of a message. The other measures may only require that a listener attend to some other details (e.g., speech rate) to form judgements on the speech samples. Although our study does support previous findings that pauses located at clause boundaries are less deleterious than pauses located clause internally, it appears that depending on the purpose of listening, it may not matter where a pause is located, as long as the pause type is facilitative.

NS listeners willingly made social judgements on NNS speech and were affected by pausing conditions in a manner that reflected responses in terms of the speech measures (fluency and comprehensibility). Both IQ and SES ratings were negatively affected by filled pauses, regardless of location. Perceived fluency and comprehensibility of speech samples were also very strongly

correlated with judgements of the speaker's IQ and SES. This seems to suggest that judgments about these social dimensions are really just based on the perceived fluency of the speaker, which may be a reasonable if faulty basis for such judgments. Whatever the explanation, evidence from this study supports previous research, which indicates that NS listeners can form negative judgments of L2 speakers on the basis of features in their speech. The hesitation free samples which would, apart from accentedness, most closely represent NS speech, were the most highly rated for both speech and social measures. As pausing may be more often utilized in NNS' speech (especially those of lower proficiency), the likelihood of negative listener evaluation increases. This study is limited in its ability to determine whether NS listeners are affected by pause type and location in similar ways when NS samples are played, but in future studies, NS controls could help determine the extent to which these negative judgements are due to the accented speech itself, or the features found in accented speech, such as pause type and pause location. It would also be beneficial to explore whether pauses located at different clause boundaries (e.g. NP, VP) are perceived similarly. Since some raters expressed feeling uncomfortable making social judgements, using a follow up questionnaire may have been beneficial to understanding why participants made the social judgements that they did, especially considering that despite the reported discomfort, there was still a clearly reliable relationship between pausing phenomena and social evaluation.

CONCLUSION

The findings of this exploratory study both support and contradict previous research. This suggests that further research is necessary to better understand the complexity in the use of pausing strategies by NNSs. Conducting a more controlled experiment where speech samples are further manipulated to determine whether pause duration impacts listener judgments, and where native speaker samples are included will lead to more conclusive results. It would also be beneficial to examine the different effects of lexical (e.g. like) versus non-lexical filled pauses since the current study was only limited to non-lexical filled pauses.

Though the results require further investigation, there may be some preliminary application for our findings. Since unfilled pauses appear to be less detrimental to listener judgements of fluency, comprehensibility, intelligence and socio-economic status, there may be some value in encouraging L2 learners to use brief silent pauses rather than filled pauses during online planning for speech production. The importance of teaching pausal phenomena extends beyond direct communication and may lead to adverse social judgements on the speakers themselves. L2 learners may be empowered if given the tools to communicate in ways that are accepted by their L2 community, while also raising their awareness that their interlocutors may make arbitrary social judgements, based on the type of pausing strategies they use. Additionally, the wider community of listeners may also benefit from knowledge that they may be making negative social judgements on the basis of features of the L2 speech signal from which the social and personal attributes of the speaker cannot actually be deduced (Munro, 2003).

REFERENCES

- Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, 49, 183-200.

- Blau, E. K. (1991). *More on comprehensible input: The effect of pauses and hesitation markers on listening comprehension*. Paper presented at the Annual Meeting of the Puerto Rico Teachers of English to Speakers of Other Languages, San Juan, Puerto Rico. Retrieved from <https://files.eric.ed.gov/fulltext/ED340234.pdf>.
- Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*(2), 159-175.
- Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014a). The perception of fluency in native and nonnative speech. *Language Learning, 64*(3), 579-614.
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014b). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language, 75*, 104-116.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274-296.
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383-398.
- Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *International Journal of Applied Linguistics, 127*(1), 53-69.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111. doi:10.1016/S0010-0277(02)00017-3.
- Davila, A., Bohara, A. K., & Saenz, R. (1993). Accent penalties and the earnings of Mexican Americans. *Social Science Quarterly, 74*(4), 902-916.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29*, 359-380.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29*(2), 320-326.
- Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language & Speech, 4*(1), 18-26.
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*(2), 135-159.

- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301-315. doi:10.1016/j.system.2010.01.005
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39(3), 569-591.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387-417.
- Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *WORD*, 15(1), 19-44.
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20(2), 38-51.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97. doi:10.1111/j.1467-1770.1995.tb00963.
- Riggenbach, H. (1991). Toward understanding fluency: A micro-analysis of nonnative speaker conversation. *Discourse Processes*, 14, 423-441.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 97-111. doi:10.1515/iral-2016-9992.
- Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, 50, 81-94. doi:10.1016/j.specom.2007.06.002.