

RESEARCH ARTICLE  

Effects of speaking task and proficiency on the midclause pausing characteristics of L1 and L2 speech from the same speakers

Amanda Huensch 

University of Pittsburgh, PA, USA

Email: amanda.huensch@pitt.edu

(Received 26 April 2022; Revised 18 April 2023; Accepted 31 May 2023)

Abstract

This study explored the effect of speaking task on midclause pausing characteristics in the L1 and L2 speech of the same speakers to gain further insights into the potential relationship between pause location and stages of speech production. Participants included English L1 learners of L2 French ($n = 29$) or Spanish ($n = 27$) from the publicly available, longitudinal LANGSNAP corpus. Participants completed two oral tasks in their L1 and L2: a picture-based narrative and a semistructured interview. The rate, duration, and proportion of midclause pauses were compared between tasks in the L1 as well as in the L2 before and during residence abroad. In the L1, results indicated more fluent performance in the narrative task except for rate. When speaking in their L2, participants showed improvement on each measure in the narrative task but ultimately remained less fluent in their L2 in comparison to their L1. In the interview task, the only measure of midclause pausing that consistently differentiated L1 from L2 speech was midclause pause rate. The findings call for a nuanced interpretation of connections between midclause pausing and formulation and suggest that midclause pause rate is least influenced by speaking task.

Introduction

An important goal of research exploring second language (L2) fluency is to better understand processes of L2 speech production. A growing body of research has indicated that pause location, as opposed to overall pause frequency or pause duration, is particularly informative when comparing L2 fluency across different proficiency levels or when differentiating between first language (L1) and L2 speech (Davies, 2003; De Jong, 2016; Huensch & Tracy-Ventura, 2017b; Kahng, 2014, 2018; Pawley & Syder, 2000; Skehan et al., 2016). Comparing L2 learners with native speakers, studies have demonstrated that although both groups have similar pausing characteristics *at* clause/message boundaries, L2 learners typically pause more often (and for longer durations) *within* clause/message boundaries (De Jong, 2016; Kahng, 2014; Skehan & Foster, 2012; Tavakoli, 2011) thereby reflecting learners' difficulties with formulation (e.g.,

grammatical and lexical encoding). De Jong (2016) reported similar results from a cross-sectional comparison of L2 speakers at different proficiency levels, and Kahng (2018) and Suzuki and Kormos (2020) demonstrated that perceived fluency ratings are sensitive to pause location.

Although the aforementioned studies have resulted in important steps forward in conceptualizing L2 speech production, much of this work has relied on findings from similar types of speaking tasks: monologic picture/video narratives (Skehan et al., 2016; Tavakoli, 2011) and responses to computer-delivered questions (De Jong, 2016; Kahng, 2014). In order to gain a more complete picture of the effects of proficiency and native-speaker status on midclause pausing and its potential relationship to stages in L2 speech production, it is necessary to expand the speaking tasks under investigation especially given a substantial body of literature (Foster & Skehan, 1996; Michel, 2011) that has demonstrated task effects on L2 fluency. At the same time, understanding how changes in task are borne out in L1 speech is also beneficial to elucidate utterance fluency characteristics that differ as a result of processing from those that differ as a result of the task (Foster & Tavakoli, 2009). In addition, the body of work evidencing differences in midclause pausing at different proficiency levels has relied on cross-sectional designs by comparing different groups of learners (De Jong, 2016; Kahng, 2014). Given the potential individual differences inherent in one's fluency characteristics (De Jong et al., 2015; De Jong & Mora, 2019; Derwing et al., 2009; Huensch & Tracy-Ventura, 2017b; Peltonen, 2018), it is desirable to compare L1 and L2 data and L2 data over time from the same speakers.

The LANGSNAP corpus (Mitchell et al., 2017) provides an ideal data set to explore the effects of task and proficiency on midclause pausing because it tracked L2 development over time using two speaking tasks and contains L1 and L2 data from the same speakers on both tasks. LANGSNAP participants included learners of L2 French ($n = 29$) or Spanish ($n = 27$) majoring in foreign languages in the UK who were required to spend their third year of university residing abroad in a French- or Spanish-speaking country. Findings have the potential to contribute to a better understanding of the effect of task on L1 and L2 speech, including L2 speech over time.

Utterance fluency and models of L2 speech production

Models of speech production provide an important framework for understanding L2 fluency and its development, and in turn, better understanding of how L2 fluency develops across different tasks can inform conceptualizations of speech production models. In the model stemming from the work of De Bot (1992) and Levelt (1989, 1999), which was further elaborated in Kormos (2006) and Segalowitz (2010, 2016), speech production consists of three main stages: *Conceptualization* involves the formation of preverbal messages; *formulation* involves grammatical, lexical, and phonological encoding; and *articulation* involves converting the phonetic plan into actual speech. An additional component of the model is *monitoring* in which speech (both planned and uttered) is checked for accuracy and appropriateness (Kormos, 2006). At multiple points in these stages are potential areas of difficulty, or "fluency vulnerability points" (Segalowitz, 2010, p. 17). Segalowitz (2016) refers to the "fluid operation (speed, efficiency) of the cognitive processes responsible for performing speech acts" (p. 82) as cognitive fluency and the measurement of temporal aspects of speech as utterance fluency. Features of utterance fluency are thus hypothesized to reflect aspects of cognitive fluency. Therefore, if L1 and L2 speakers differ in their linguistic knowledge

and access to it and the goal is to better understand L2 speakers' knowledge and access, one approach is to make comparisons between groups and interpret positive changes as reflecting knowledge gains or improvements in access. For instance, investigations could include comparing which utterance fluency measures change over time for L2 speakers or which utterance fluency measures differ among L2 speakers of different proficiency levels, or even which utterance fluency measures differ between L1 and L2 speakers.

Connecting utterance fluency measurements to cognitive fluency or stages in speech production, however, is not a straightforward endeavor. In conceptualizing how to characterize utterance fluency subdimensions, Tavakoli and Skehan (2005) categorized previous operationalizations of fluency into three subdimensions: speed fluency (the rate of speech), breakdown fluency (silent and filled pauses), and repair fluency (reformulations). More recently, Skehan et al. (2016) called for a reconceptualization of how utterance fluency is characterized, and how it is connected to models of L2 speech production. Instead of speed, breakdown, and repair, they argued for a two-way distinction: discourse fluency and clause fluency. Discourse fluency is connected to the conceptualization stage of the L2 speech production model and entails developing ideas and connecting larger discourse units. Thus, it is concerned with macro-planning and any disfluency issues will occur at clause or utterance boundaries. Clause-level fluency is connected to formulation, which includes processes such as lexical retrieval and syntactic encoding, any disfluency issues will occur within clauses. Skehan et al. suggested that this two-way distinction is preferable because although both L1 and L2 speakers must pause, the location of their pauses would likely differ because L1 speakers pause more for conceptualization (i.e., at clause boundaries) whereas L2 speakers pause more for formulation. The same could also be argued for L2 speakers at different proficiency levels such that lower proficiency learners would be predicted to pause more within clause boundaries than learners at higher proficiencies. In this conceptualization of utterance fluency and its connections to models of L2 speech production, midclause pauses are taken as evidence of formulation difficulties.

Several studies have directly examined the relationship between cognitive and utterance fluency by including measures of both aspects of fluency in their study design (De Jong et al., 2013; Kahng, 2020; Segalowitz & Freed, 2004; Suzuki & Kormos, 2023). For instance, Kahng (2020) was interested in exploring not only the relationship between utterance and cognitive fluency but also the role of L1 utterance fluency. She did so by relating a number of L1 and L2 utterance fluency measures taken from performance on monologic speaking tasks to the results of a battery of cognitive fluency measures and found that although most L2 utterance fluency measures appeared to be connected to multiple cognitive fluency measures and the corresponding L1 fluency measure, the rate of midclause pauses was unique in that it was the only utterance fluency feature that was predicted solely by an L2 cognitive fluency measure. Similarly, Suzuki and Kormos (2023), using two monologic and two read-aloud tasks, reported that a measure of midclause pause frequency contributed significantly more to predicting breakdown fluency in a structural equation model than did any other utterance fluency measure of breakdown fluency included (i.e., length and frequency of pauses at clause boundaries and a measure of filled pausing). Because the midclause pause measure was consistent across the monologic and read-aloud tasks, it was suggested that it could be a potentially strong candidate for use in automatic scoring of oral proficiency.

Finally, Segalowitz (2016) highlighted the necessity of situating investigations of L2 fluency (both cognitive and utterance) within their social context. Drawing on usage-

based approaches to understanding language acquisition and communication, Segalowitz argued that “normal communication involves interlocutors attempting to establish joint attention and reading each other’s social intentions” (p. 88) and in combination with transfer appropriate processing (i.e., how memories are retrieved is related to how they were encoded) this supports developing L2 fluency in contexts involving attentional/intentional demands. Similarly, this conceptualization has implications for how speech data are collected: The inclusion or not of having to handle joint attention and infer social intentions might affect utterance fluency. Arguably in monologic narrative tasks there are lower demands on a speaker in terms of both joint attention and inferring social intentions in comparison with participating in a semistructured interview. For example, we know from research investigating dialogic contexts, speakers must manage aspects of turn-taking and interaction, which Peltonen (2017) referred to as dialogic fluency, including an added time pressure to plan as well as the necessity of responding at appropriate points (Garrod, 1999). Regarding the latter, van Os et al. (2020) examined perceptions of fluency in dialogic speech and demonstrated that experimentally manipulated turn-taking behaviors had an influence on how raters judged fluency. It is also the case that studies comparing utterance fluency in monologic versus dialogic tasks often indicate higher fluency in dialogues (Sato, 2014; Tavakoli, 2016). In summary, following the line of argumentation in Segalowitz (2016) and what we know about how fluency might differ in monologic versus dialogic contexts, it is necessary to expand the investigation of midclause pausing beyond monologic tasks.

Pause location and task effects in L1 and L2 fluency

The finding that pause location, as opposed to overall frequency or duration, differentiates L1 from L2 speech as well as L2 speech at different proficiency levels has been demonstrated in a handful of studies using monologic picture/video narratives and responses to computer-delivered open-ended questions (De Jong, 2016; Foster & Tavakoli, 2009; Kahng, 2014; Skehan et al., 2016; Tavakoli, 2011). An initial driving force to investigate pause location in L2 speech stemmed from L1 literature (e.g., Goldman Eisler, 1972; Pawley & Syder, 2000) that provided some evidence that pauses in L1 speech tend to occur more often at/near clause boundaries than between them. In comparison with L1 speakers, it is hypothesized that L2 speakers will pause more often within clauses because they most likely do not have as substantial a lexicon and/or efficient access to it (Kormos, 2006; Skehan et al., 2016). In her investigation, Kahng (2014) compared the pausing characteristics of L1 and L2 speech from different speakers who completed a computer-delivered task in which they were prompted to speak for 1 min each about their field of study and free-time activities. Her results indicated that although silent pause duration and filled pause usage patterns did not clearly differentiate L1 from L2 speech, the rate of silent pauses within a clause for L2 speakers was twice that of L1 speakers, and this measure negatively correlated with L2 proficiency such that the higher the proficiency the lower the rate of midclause pausing.

Similarly, in another cross-sectional study De Jong (2016) demonstrated that L1 and L2 speakers differ with respect to pause location in her investigation of Turkish and English learners of Dutch. De Jong differentiated pausing that occurred within and between analysis of speech units (ASU; Foster et al., 2000) as opposed to within and between clauses, but she importantly pointed out that taking ASU length into consideration was necessary to avoid potential confounds between longer utterances and a higher likelihood to pause. Kahng (2018) also accounted for clause length in

her normalization of utterance fluency measures. Both Kahng and De Jong argued that their findings of the importance of clause location provide implications for language assessment tools such that more valid measures of L2 fluency ought to incorporate the aspects of utterance fluency that have been demonstrated to differentiate L1 from L2 speech.

A small set of studies has investigated potential task effects on pause location in L1 and L2 speech (Foster & Tavakoli, 2009; Skehan et al., 2016; Tavakoli, 2011) with a specific focus on understanding how different aspects of narrative tasks might affect fluency. One aspect of narrative tasks that has been investigated is tight versus loose structure (Foster & Tavakoli, 2009; Tavakoli & Foster, 2008), or in other words, whether the temporal order of the storyline must be presented in a certain sequence for it to make sense. For instance, Foster and Tavakoli (2009) compared the effects of narrative structure on L1 speaker speech and compared it with their L2 data from Tavakoli and Foster (2008). Their results indicated that although narrative structure did not appear to affect L1 fluency, tightly structured narratives had a positive, albeit modest effect on L2 performance. As also demonstrated in Tavakoli (2011), findings indicated that native speakers paused less frequently at midclause locations in comparison with nonnative speakers. In their discussion, they called for an exploration that compares learners in their L1 on multiple tasks in addition to completing comparisons of those same learners' L2s, the focus of the current study.

Another aspect of a narrative task that has been demonstrated to affect L2 fluency is related to the necessity of including certain lexical items or structures to successfully retell the story (Derwing et al., 2004; Skehan & Foster, 2012). Although not focused on comparing L1–L2 speech, Derwing et al. (2004) compared perceived fluency ratings of L2 speech across three different tasks, including a picture narrative, and provided evidence that the lowest ratings of perceived fluency were found on the narrative task. They hypothesized that task differences “may reflect task-dependent variability in the degree of freedom the speaker had in choosing lexical items, structures, and content in general” (pp. 670–671). Similarly, Skehan and Foster (2012) reported that having to include necessary elements in a task appeared to negatively affect L2 fluency but did not affect L1 fluency in the same way. In other words, the L1–L2 fluency differences reported for midclause pausing in previous studies might be particularly pronounced because of the use of narrative tasks.

Bringing together the findings from previous work, midclause pausing appears to be a relatively robust utterance fluency measure that differentiates L1 from L2 speech. Nevertheless, these findings have heavily relied on investigations using monologic, narrative tasks, whereas Segalowitz (2016) has called for expanding our understanding of L2 fluency to include contexts involving attentional/intentional demands such as an interview task. Finally, Foster and Tavakoli (2009) and Tavakoli and Foster (2008) argued that having L1 speaker baseline data is necessary to investigate differences across tasks to make claims about differences in L1 versus L2 speech-production processes. Therefore, the current analysis explores the midclause pausing of L1 and L2 speech (from the same speakers) in a picture-based narrative task and a semistructured interview task.

Current study

Framed by previous research using monologic tasks that has found differences in midclause pausing rates between L1 and L2 speakers and L2 speakers at different

proficiency levels but used speech from different speakers, the current study compared the rate, duration, and proportion of midclause silent pauses in a picture-based narrative and a semistructured interview using the LANGSNAP corpus. The LANGSNAP corpus has been used previously for investigations of fluency development (Huensch & Tracy-Ventura, 2017a, 2017b) and maintenance postinstruction (Huensch et al., 2019). For instance, Huensch and Tracy-Ventura (2017b) examined the speed, breakdown, and repair fluency of the Spanish subset across the six data collection waves before, during, and after study abroad and demonstrated that those elements of utterance fluency that improved quickly were those that were maintained even after being back home in the L1 environment for 8 months. In each of these three prior studies exploring fluency in the LANGSNAP corpus, the only task reported on was the picture narrative. Additionally, none of those studies incorporated a fluency measure of midclause pauses. Thus, the current study provides a unique contribution by using data from the oral interview task and focusing on new measures of utterance fluency: midclause silent pause rate, duration, and proportion. The LANGSNAP corpus is ideal to investigate the questions of the current study because it includes L1 and L2 speech from the same speakers and L2 speech from the same speakers at two points before and during study abroad where proficiency (as measured by an elicited imitation test) increased.¹ By investigating L1 speakers' pausing behavior across a wider range of speaking styles, we can gain further insights into the potential relationship between pause location and stages of speech production.

Research questions

1. To what extent are the rate, duration, and proportion of midclause silent pauses in L1 speech similar across a narrative and interview task?
2. To what extent are the rate, duration, and proportion of midclause silent pauses in the L1 and L2 speech of the same speakers similar within a narrative task and an interview task as proficiency increases in the L2?

Method

Study design

Data in the current study are a subset of the publicly available corpus of a 2-year longitudinal project investigating university students' language development during and after study abroad: the Languages and Social Networks Abroad Project (LANGSNAP; Mitchell et al., 2017). LANGSNAP included both learners of French and Spanish, and data were collected once before (Pre-sojourn), three times during (In-sojourn 1, In-sojourn 2, and In-sojourn 3), and twice after (Postsojourn 1 and Postsojourn 2) students resided abroad. The LANGSNAP data are ideal for answering the research questions in the current study because they allow for a within-subjects comparison of L2 data over a period of demonstrated improvement in proficiency as well as a comparison of L1 and L2 data from the same speakers, with two different oral tasks available for all comparisons (a picture-based narrative and a semistructured

¹As reported in Huensch and Tracy-Ventura (2017b, p. 766), both the French and Spanish learners demonstrated significant improvements in their proficiency. Effect sizes indicated medium ($d = 1.12$ for the French group) and large ($d = 1.51$ for the Spanish group) effects in relation to changes in elicited imitation test scores. These results demonstrate evidence of proficiency changes for these participants.

interview, described in the Materials and Procedure section). L1 data were collected twice: the interview at In-sojourn 3 and the narrative at Postsojourn 2. The point of data collection was not considered in the analysis of the L1 data given the assumption that L1 fluency in this population (adult, instructed L2 learners) would be relatively stable over time, particularly in comparison with L2 fluency, as linguistic knowledge and access to it is likely more robust and efficient in the L1 (see also the Discussion section). The L2 narrative and interview data in the current analysis are from the Presojourn and In-sojourn 2 (approximately 5 months into the learners' stay abroad). Participants completed both tasks in the L2 at each point. The Presojourn and In-sojourn 2 data points were chosen because at those points, and not during In-sojourn 1 or In-sojourn 3, a proficiency test was administered in the form of an elicited imitation test (EIT; Bowden, 2016; Ortega, 2000). Thus, it is possible to demonstrate that participants' proficiency improved between these points. Two points, and importantly two points between which participants' L2 proficiency improved during study abroad, were compared for the L2 data to determine whether midclause pausing behaviors changed as proficiency increased. In-sojourn 2 was selected rather than Postsojourn 1 because participants were still immersed in the target language environment.

Participants

The LANGSNAP participants were 56 undergraduate students who spent their third year of a 4-year degree living abroad in a French-speaking ($n = 29$) or Spanish-speaking ($n = 27$) country. All participants were majoring in modern languages and were paid for their participation. Most participants reported studying a language other than French or Spanish either before or during university. This information and further details about the project and participants (including the publicly available data) can be found at the LANGSNAP web site: <http://langsnap.soton.ac.uk>. Some participants' data were excluded from the current analysis because either English was not their L1 or there were missing or low-quality sound files (participants 100, 108, 122, 126, 150, 158, and 165). Table 1 summarizes the age, prior years of L2 instruction, and EIT results (demonstrating increased oral proficiency for both groups with medium to large effects) for the 49 participants in the current study separated by L2 group.

Materials and procedure

Oral data in the LANGSNAP corpus include productions from two types of tasks: (a) picture-based narratives and (b) semistructured interviews. Two versions of the narrative task (both available on IRIS; <https://www.iris-database.org/iris/>) are included in the current analysis, the Cat Story (Presojourn in L2 and L1; Domínguez et al., 2013)

Table 1. Participant demographic and proficiency information

	Spanish L2 Learners ($n = 24$)		French L2 Learners ($n = 25$)	
Number of participants	24 (17 females)		25 (23 females)	
Mean age (SD)	21 (1.2)		20 (0.7)	
Years studying L2 (SD)	6 (3.2)		11 (2.4)	
EIT (out of 120) (SD)	Presojourn	In-sojourn 2	Presojourn	In-sojourn 2
	84 (12)	99 (8)	59 (16)	77 (17)
	$t(23) = 7.97, p < .001, d = 1.51$		$t(24) = 9.18, p < .001, d = 1.12$	

and the Brothers Story (In-sojourn 2 in L2; based on the children's story *I Very Really Miss You*; Langley, 2006). Both stories were approximately 15 pages in length and included prompts in either the L1 (English) or the L2 (French or Spanish, e.g., *La historia de Natalia y su gato Pancho/L'histoire de Natalie et de son chat Pompon/The story of Natalia and her cat Pancho*). Multiple narratives were used in LANGSNAP to avoid repetition effects across the six data collection points; however, the narratives were designed and piloted to be as similar as possible. Participants were given a few minutes to look at the pictures to gain a general idea of the plot line of the story. After that time, they were asked to retell the story in their own words while continuing to be able to look at the pictures. No time limit was imposed on the responses; thus, responses varied somewhat but were similar overall in length (see Table 2). Importantly, the procedure across the narrative and interview tasks was parallel in that neither included a time limit.

Interview data were collected via a semistructured interview with approximately 10 questions that participants completed with a member of the project team. The questions focused on topics related to the participants' opinions and experiences related to their time abroad or their hopes/expectations for their time abroad at the pretest. The interviews lasted approximately 10–20 min each, and interviewers were instructed not to offer help (e.g., lexical item, verb conjugation) such that participants could be allowed to say as much as they could on their own. However, if the participants explicitly requested assistance, the interviewers were instructed to provide it. Although the interviewers were instructed to allow the participants to say as much as they could on their own, they were also advised to be active listeners: demonstrating signs of understanding by nodding, smiling, etc. To be able to compare similar amounts of speech between the interview and narrative tasks, for the purpose of the current analysis, only a portion of the interview data was analyzed: Participants' responses to the first question and a question approximately halfway through the interview were used (see the Appendix for the specific questions used from each data collection point). Table 2 provides the means and standard deviations of the duration of speech for each of the tasks at the Presojourn, In-sojourn 2, and in the L1 English. The total duration of the oral-production data in the current study is 11 hr and 24 min.

As a final consideration, it is important to note that although using existing, publicly available corpora has multiple benefits, including broadening the utility of the data collected (MacWhinney, 2017; Tracy-Ventura & Huensch, 2018), there can also be potential methodological limitations—for example, in the current study not tightly controlling task design features via manipulation. To address this potential limitation, measures of lexical and syntactic complexity were also calculated and incorporated into the analysis with the objective of controlling for the effects of any potential differences

Table 2. Mean duration (and standard deviation) of speech samples from the narrative and interview tasks

	Presojourn (L2)		In-sojourn 2 (L2)		English (L1)	
	Narrative	Interview	Narrative	Interview	Narrative	Interview
Spanish L2 learners (<i>n</i> = 24)	263 s (107 s)	80 s (25 s)	185 s (87 s)	103 s (48 s)	108 s (53 s)	88 s (29 s)
French L2 learners (<i>n</i> = 25)	298 s (107 s)	80 s (41 s)	176 s (71 s)	113 s (75 s)	115 s (36 s)	81 s (37 s)

when examining the main research question of task effects on midclause pausing behavior.

Data coding

Data were transcribed in CLAN following CHAT conventions (MacWhinney, 2000) and separated into ASUs (Foster et al., 2000). Foster et al. defined ASUs as “consisting of an independent clause, or sub-clausal unit, together with any subordinate clause (s) associated with either” (emphasis in original) (p. 365). Each transcript, including ASU placement, was checked by at least two members of the research team. In order to conduct an investigation of midclause pausing, it was necessary to mark clauses in the transcripts. This was done using the code ‘[^c]’. Clauses were defined as consisting “minimally of a finite or non-finite Verb element plus at least one other clause element (Subject, Object, Complement or Adverbial)” (Foster et al., 2000, p. 366). Two coders independently coded clauses in a subset of the data. Interrater reliability comparing the number of clauses coded reached acceptable levels (Cronbach’s alpha = .99).

Next, instances of speech and silence were automatically segmented in Praat (Boersma & Weenik, 2015) using the *Annotate To TextGrid (silences...)* command after which each TextGrid was manually checked. This step was completed to catch any inaccuracies of the automatic segmentation program (e.g., a cough being identified as a speech segment). Minimum silent pause duration was set at 250 ms (De Jong & Bosker, 2013). Next, the transcripts were exported as TextGrids and merged with the existing speech/silence TextGrids. The transcript coding was then used to code all silent pauses as either (1) within a clause, (2) at a clause boundary, or (3) at an ASU boundary. After a round of training and discussion, two coders independently coded a subset of the data. The codes were compared, and interrater reliability reached acceptable levels (Cronbach’s alpha = .99). Finally, a Praat script was used to automatically tabulate the number and duration of the pause and speech segments.

Three measurements of midclause silent pausing were calculated representing (a) the rate (or frequency) of midclause silent pauses, (b) the duration of midclause silent pauses, and (c) the proportion of midclause to end-clause silent pauses. Rate, following Kahng (2018), was calculated by dividing the total number of midclause silent pauses by the number of clauses and the number of words per clause (number of midclause pauses/number of clauses/number of words per clause). This measure represents “on average how often a speaker pauses within a clause ... normalized per word to take into account length of clauses” (Kahng, 2018, p. 576). Duration was calculated by dividing the total duration of midclause silent pauses (in ms) by the total number of midclause silent pauses. This measure is thus the average length of midclause silent pauses (in ms). Finally, for the proportion of midclause pausing, the duration of midclause silent pauses was divided by the duration of all silent pauses. Thus, a proportion of .50 would mean that half of the silent pause duration occurred within a clause and half at a clause or ASU boundary, a proportion above .50 would indicate a higher proportion of silent pause duration at midclause than at end clause, and a proportion below .50 would indicate a lower proportion of silent pause duration at midclause than at end clause. This variable was normalized to take into account the length of clauses by dividing the proportion by the number of words per clause.

Finally, measures of lexical and syntactic complexity were calculated to control for any potential differences across the two tasks. Lexical complexity was operationalized as lexical diversity (Jarvis, 2013) and computed using the MATTR command on the POS

tagged transcripts in CLAN with a window length of 10 words. MATTR was selected as it has been shown to be less sensitive to text length (Fergadiotis et al., 2015). For syntactic complexity, a commonly used measure was calculated to represent subordination: the ratio of clauses to ASUs (De Clercq & Housen, 2017). Clause and ASU counts were extracted from the transcript using CLAN `FREQ` commands.

Analysis

For all analyses, linear mixed-effects models were calculated in R (Version 4.2.2; R Core Team, 2022) using the *lme4* package (Bates et al., 2014); data and the R code are available at <https://osf.io/dn6v3/>. Separate models were fit for each of the midclause pause variables: rate, duration, and proportion. Final model structures were determined using backward elimination via the *lmerTest* package *step* function (Kuznetsova et al., 2017), which computes *p* values using Satterthwaite's degrees of freedom method. Descriptive statistics and graphs (box plots, histograms, and QQplots) for each of the midclause pause measures are provided in the supplementary materials along with plots corresponding to the final models checking for linearity, homogeneity of variances, and normally distributed model residuals. The purpose of research question one was to examine the extent to which midclause silent pause rate, duration, and proportion are similar in the L1 across the narrative and interview tasks. Therefore, each initial model began with all fixed effects of potential interest: task (narrative, interview), lexical complexity (MATTR score) and syntactic complexity (clause/ASU), and random intercepts for participant.

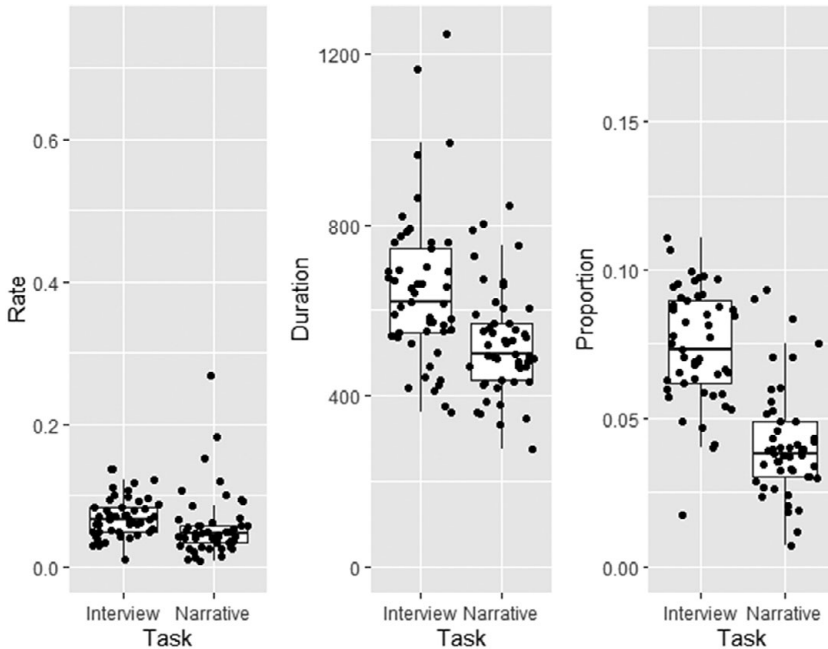
The focus of research question two was to examine whether the L1-L2 patterning of midclause silent pauses was similar in the narrative and interview tasks. Each initial model began with all fixed effects of potential interest: task (narrative, interview), L2 group (French, Spanish), round (L2 at Presojourn, L2 at In-sojourn 2, and L1), lexical complexity (MATTR score) and syntactic complexity (clause/ASU), and random intercepts for participant. L2 group was included as a fixed effect in case of any potential cross-language differences in midclause pausing. Marginal and conditional R^2 values are reported for the final models and interpreted following Plonsky and Ghanbar's (2018) recommendations of values lower than 0.20 representing small effects and values greater than 0.50 representing large effects. Estimated marginal mean (emm) values were plotted to allow for interpretation of each of the final models. The supplementary materials contain the estimated marginal mean values and 95% CIs for each final model as well as comparisons between the maximal and final models for each measure for research question two.

Results

For research question one, first descriptive statistics and box plots of the three measurements of midclause silent pausing (rate, duration, and proportion) are presented followed by the results of the mixed-effects model analyses. In the box plot figures, each dot represents an individual data point. The first research question investigated the extent to which rate, duration, and proportion of midclause silent pauses in L1 speech were similar across the narrative and interview tasks. Table 3 provides the means, standard deviations, and corresponding 95% CIs for each of the midclause pause measures for the L1 in the narrative and interview tasks. Figure 1 displays the box plots of midclause pause rate, duration, and proportion in the L1 for

Table 3. Means (SDs) and corresponding 95% confidence intervals of midclause silent pause rate, duration, and proportion in the interview and narrative tasks in the L1

	Rate	Duration	Proportion
Narrative	0.057 (0.046) [0.044, 0.070]	523.35 (124.33) [487.64, 559.06]	.041 (.019) [.035, .046]
Interview	0.070 (0.028) [0.062, 0.078]	648.30 (182.55) [595.87, 700.74]	.074 (.019) [.069, .080]

**Figure 1.** Box plots of midclause silent pause rate, duration, and proportion in the L1 for the interview and narrative tasks.

the interview and narrative tasks. As seen in Figure 1, midclause pausing in L1 speech appears to be influenced by task such that the rate, duration, and proportion of midclause silent pauses are lower in the narrative task. Visually, this difference appears largest for the proportion measure and smallest for the rate measure, which is further supported by the CIs reported in Table 3. The nonoverlapping CIs for the duration and proportion measures suggest differences between the narrative and interview tasks, whereas the overlapping CIs for the rate measure (upper limit for the narrative is 0.070 compared with the lower limit for the interview 0.062) indicate no difference.

Tables 4 and 5 report the final models for rate, duration, and proportion. The results indicated a significant difference between the narrative and interview task in the L1 for the midclause pause measures of duration and proportion but not rate. For rate, as seen in Table 4, the final model had a marginal R^2 value of only .06 and 95% CIs crossing through 0, indicating a negligible effect. For duration (Table 5), the only significant fixed effect in the final model structure was task, $\beta = -124.95$, $SE = 28.60$, 95% CI $[-181.74, -68.17]$, $p < .001$, with the model indicating that when speaking in the L1, the

Table 4. Summary of mixed-effects model fit for L1 rate

Fixed effects	Maximal model					Final model				
	Estimates	SE	95% CI	<i>t</i>	<i>p</i>	Estimates	SE	95% CI	<i>t</i>	<i>p</i>
(Intercept)	−0.24	0.23	[−0.70, 0.22]	−1.04	.299	0.02	0.02	[−0.01, 0.05]	1.28	.202
Task [narrative]	−0.01	0.01	[−0.02, 0.00]	−1.31	.192					
MATTR	0.30	0.25	[−0.20, 0.79]	1.19	.238					
Syntactic complexity	0.02	0.01	[−0.00, 0.03]	1.88	.064	0.02	0.01	[0.00, 0.03]	2.62	.010
Random effects										
σ^2	0.00					0.00				
τ_{00}	0.00	Participant				0.00	Participant			
ICC	0.31					0.31				
<i>N</i>	49	Participant				49	Participant			
Observations	98					98				
Marginal R^2 / Conditional R^2	.076 / .363					.059 / .354				

duration of midclause silent pauses was approximately 125ms shorter in the narrative task. The marginal R^2 value (.14) indicated a small effect. Finally, the results indicated that when speaking in the L1, the proportion of midclause silent pauses is lower in the narrative task. Specifically, when considering the relative amount of time spent pausing within and between clauses, the proportion of the time spent pausing within a clause was almost two times larger in the interview than in the narrative task, $\beta = -0.03$, narrative $M = 0.041$, interview $M = 0.074$. The final model structure indicated significant effects of task and syntactic complexity with a marginal R^2 value (.48) indicating a medium effect.

To summarize, when controlling for lexical and syntactic complexity and comparing performance between the tasks, L1 speakers demonstrated higher levels of fluency in the narrative task for the duration and proportion measures but not the rate measure.

The second research question investigated the extent to which rate, duration, and proportion of midclause silent pauses in the L1 and L2 speech of the same speakers are similar across tasks as proficiency increases in the L2. Table 6 provides the means (standard deviations) and corresponding 95% CIs for each of the measurements of midclause silent pausing in the narrative and interview tasks for the L2 at Presojourn, L2 at In-sojourn 2, and the L1. Figure 2 displays the box plots for the rate measure.

As seen in Figure 2, a similar pattern emerged in both tasks such that the rate of midclause silent pauses appeared to decrease from L2 at Presojourn to L2 In-sojourn 2 and was lowest in the L1. As shown in Table 7, the final model had a marginal R^2 value of .50, indicating a large effect. The model did not include any simple or interaction effects for L2Group, indicating comparability across the groups. The final model did, however, include significant simple effects for both task and round, and importantly, significant interactions between task and round. Figure 3 plots the estimated marginal means with 95% CIs and demonstrates that the greatest difference between the tasks in terms of the rate of midclause silent pauses occurs in the L2 at Presojourn—0.195 [0.172, 0.219] vs. 0.291 [0.267, 0.315]—whereas rate is comparable in both tasks in the L1—0.070 [0.047, 0.094] vs. 0.059 [0.036, 0.082]. Figure 3 also illustrates similarity across the tasks for rate such that in both the narrative and interview tasks, learners became more fluent during their time abroad—as indicated by a decrease in midclause silent pause rates—but remained less fluent in their L2 in comparison to their L1.

Table 5. Summary of final mixed-effects model fits for L1 duration and proportion

Fixed effects	Duration: Final model					Proportion: Final model				
	Estimates	SE	95% CI	<i>t</i>	<i>p</i>	Estimates	SE	95% CI	<i>t</i>	<i>p</i>
(Intercept)	648.30	22.31	[604.00, 692.60]	29.06	<.001	0.05	0.01	[0.03, 0.06]	4.74	<.001
Task [narrative]	-124.95	28.60	[-181.74, -68.17]	-4.37	<.001	-0.03	0.00	[-0.04, -0.02]	-8.33	<.001
MATTR										
Syntactic complexity						0.01	0.00	[0.00, 0.02]	3.09	.003
Random effects										
σ^2	20,039.25					0.00				
τ_{00}	4,352.17	Participant				0.00	Participant			
ICC	0.18					0.22				
<i>N</i>	49	Participant				49	Participant			
Observations	98					98				
Marginal R^2 / Conditional R^2	.139 / .293					.476 / .590				

Table 6. Means (*SDs*) and corresponding 95% confidence intervals of midclause silent pause rate, duration, and proportion in the narrative and interview tasks for L2 at Presojourn, L2 at In-sojourn 2, and L1

	Rate		Duration		Proportion	
	Narrative	Interview	Narrative	Interview	Narrative	Interview
Presojourn	0.297 (0.131) [0.259, 0.334]	0.191 (0.087) [0.165, 0.216]	755.58 (141.12) [715.05, 796.12]	614.28 (136.95) [574.94, 653.62]	.090 (.019) [.085, .096]	.099 (.027) [.091, .107]
In-sojourn 2	0.158 (0.094) [0.131, 0.185]	0.113 (0.066) [0.094, 0.132]	619.68 (110.58) [587.92, 651.44]	606.17 (129.90) [568.86, 643.49]	.067 (.020) [.061, .071]	.077 (.023) [.071, .084]
L1	0.057 (0.046) [0.044, 0.070]	0.070 (0.028) [0.062, 0.078]	523.35 (124.33) [487.64, 559.06]	648.30 (182.55) [595.87, 700.74]	.041 (.019) [.035, .046]	.074 (.019) [.069, .080]

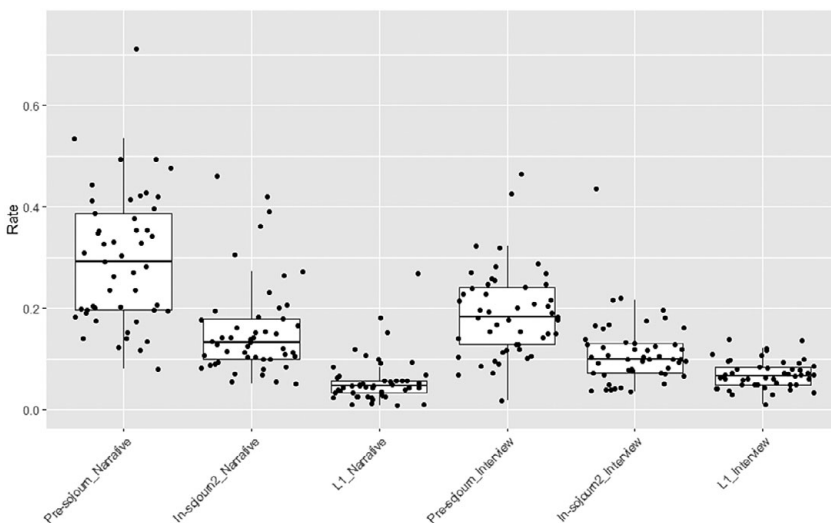


Figure 2. Box plots for rate for the narrative and interview tasks for L2 at Presojourn, L2 at In-sojourn 2, and L1.

Table 7. Summary of final mixed-effects model fits for rate

Fixed effects	Rate: Final model				
	Estimates	SE	95% CI	t	p
(Intercept)	0.61	0.21	[0.20, 1.03]	2.90	.004
Task [narrative]	0.10	0.01	[0.07, 0.12]	7.33	<.001
Round [In-sojourn_2]	-0.08	0.01	[-0.11, -0.06]	-6.74	<.001
Round [L1]	-0.12	0.01	[-0.15, -0.10]	-10.23	<.001
MATTR	-0.44	0.22	[-0.88, -0.01]	-2.00	.047
Task [narrative] × Round [In-sojourn_2]	-0.05	0.02	[-0.09, -0.02]	-2.83	.005
Task [narrative] × Round [L1]	-0.11	0.02	[-0.14, -0.07]	-5.97	<.001
Random effects					
σ^2	0.00				
τ_{00}	0.00	Participant			
ICC	0.47				
N	49	Participant			
Observations	294				
Marginal R^2 / Conditional R^2	.499 / .735				

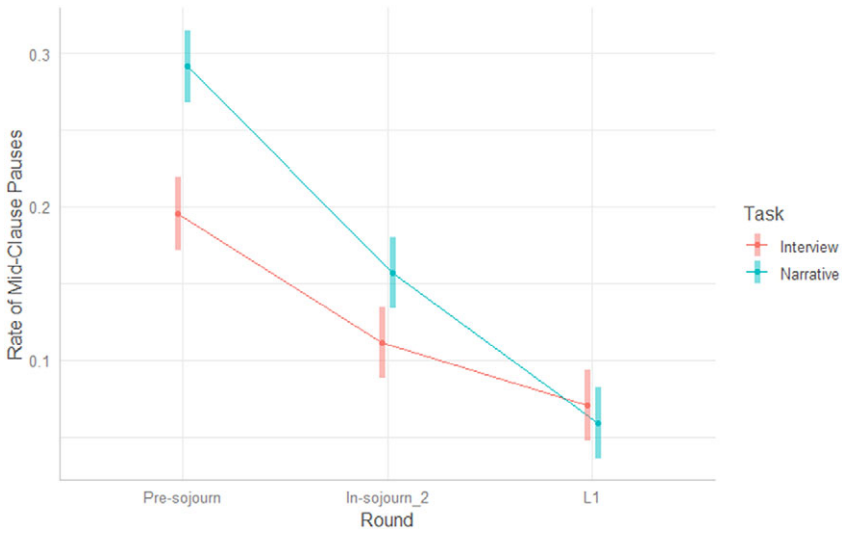


Figure 3. Estimated marginal means and 95% CIs for the final rate model.

Next the results for the duration measure are presented. Figure 4 displays the corresponding box plots for the duration measure.

As seen in Figure 4, the pattern that emerged for midclause pause duration is similar to that of rate on the narrative task (although there potentially seems to be slightly more variation in midclause pause durations): the duration of midclause silent pauses appears to decrease from L2 at Presojourn to L2 In-sojourn 2 and is lowest in the L1. However, in contrast, the duration of midclause silent pauses does not appear to differ

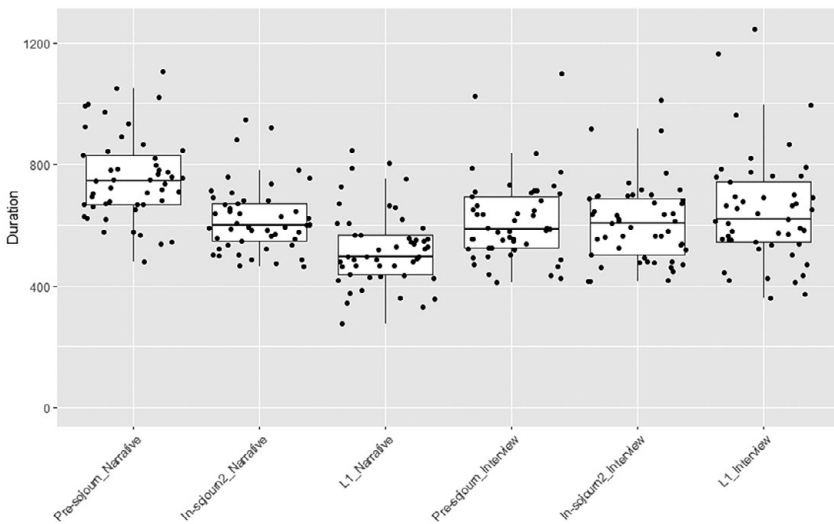


Figure 4. Box plots for duration for the narrative and interview tasks for L2 at Presojourn, L2 at In-sojourn 2, and L1.

across rounds in the interview task. The results of the mixed-effects model analysis support this: As shown in Table 8, the final model had a marginal R^2 value of .20, indicating a small effect. The model did not include any simple or interaction effects for L2Group, indicating comparability across the groups. The final model did, however, include a significant simple effect for task and statistically significant interactions between task and round.

Figure 5 plots the estimated marginal means with 95% CIs and demonstrates that, similar to the results for rate on the narrative task, the results for duration on the narrative task indicate that learners became more fluent during their time abroad—756ms [716, 795] vs. 620ms [580, 659]—but remained less fluent in their L2 in comparison

Table 8. Summary of final mixed-effects model fits for duration

Fixed effects	Duration: Final model				
	Estimates	SE	95% CI	<i>t</i>	<i>p</i>
(Intercept)	614.28	19.91	[575.09, 653.47]	30.85	<.001
L2Group [Spanish]					
Task [narrative]	141.30	22.86	[96.31, 186.30]	6.18	<.001
Round [In-sojourn_2]	-8.10	22.86	[-53.10, 36.89]	-0.35	.723
Round [L1]	34.02	22.86	[-10.97, 79.02]	1.49	.138
Task [narrative] × Round [In-sojourn_2]	-127.79	32.33	[-191.43, -64.16]	-3.95	<.001
Task [narrative] × Round [L1]	-266.26	32.33	[-329.89, -202.62]	-8.24	<.001
Random effects					
σ^2	12,804.08				
τ_{00}	6,622.12	Participant			
ICC	0.34				
<i>N</i>	49	Participant			
Observations	294				
Marginal R^2 / Conditional R^2	.196 / .470				

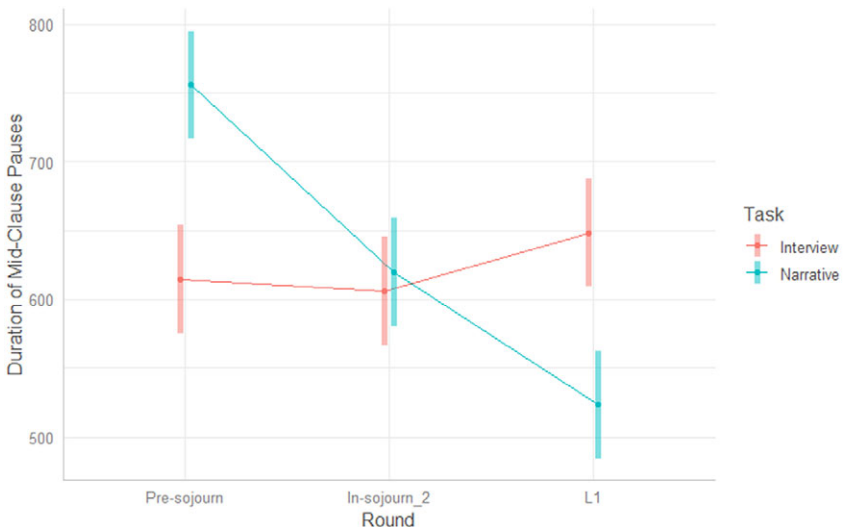


Figure 5. Estimated marginal means and 95% CIs for the final duration model.

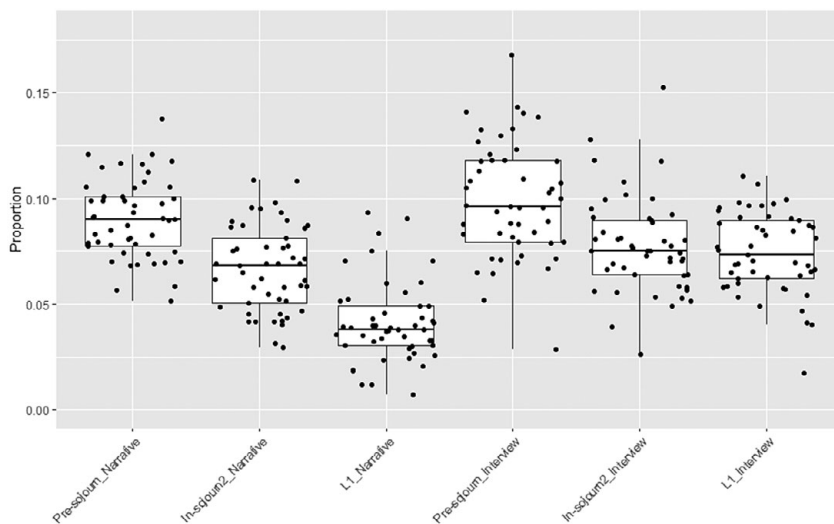


Figure 6. Box plots for proportion for the narrative and interview tasks for L2 at Presojoir, L2 at In-sojoir 2, and L1.

with their L1—523ms [484, 563]. In contrast, the figure illustrates no differences across the rounds for duration on the interview task, with largely overlapping CIs in L2 at Presojoir, [575, 654]; L2 at In-sojoir 2, [567, 645]; and L1, [609, 688].

Finally, the results for the proportion measure are presented. **Figure 6** displays the corresponding box plots for the proportion measure.

As seen in **Figure 6**, it was again the case that on the narrative task the proportion of midclause silent pauses appeared to decrease from L2 at Presojoir to L2 In-sojoir 2 and was lowest in the L1. On the interview task, it appeared that the proportion of midclause silent pauses decreased from L2 at Presojoir to L2 In-sojoir 2 but that the proportion of midclause silent pauses was similar for L2 at In-sojoir 2 and L1. As shown in **Table 9**, the final model had a marginal R^2 value of .48 indicating a large effect. Unlike the previous models, the final model included simple and interaction effects for L2Group, indicating differences between the French and Spanish learner groups. Similar to the rate and duration models, the final model included no three-way interaction.

Figure 7 plots the estimated marginal means with 95% CIs and shows the interaction of task and round separately for each L2 group. As demonstrated in the figure, across the L2 groups the results for proportion showed a trend similar to those of rate and duration on the narrative task: Although the speakers were significantly more fluent in L2 at In-sojoir 2 compared with L2 at Presojoir, they were the most fluent in L1. On the interview task, although the Spanish learners show slightly lower proportions in the L1, 0.067 [0.058, 0.076], than in the L2 at In-sojoir 2, 0.086 [0.077, 0.094], the French learners show comparable midclause silent pause proportions in L2 at In-sojoir 2 and L1—0.074 [0.066, 0.082] vs. 0.070 [0.061, 0.078]—with overlapping CIs.

To summarize, speaking task appeared to affect midclause silent pausing in both L1 and L2 speech. When speaking in their L1, participants demonstrated higher fluency on the narrative task as indicated by shorter and a lower proportion of midclause silent pauses. In terms of development over time, when speaking their L2, participants

Table 9. Summary of final mixed-effects model fits for proportion

Fixed effects	Proportion: Final model				
	Estimates	SE	95% CI	<i>t</i>	<i>p</i>
(Intercept)	0.06	0.01	[0.05, 0.08]	9.43	<.001
L2Group [Spanish]	0.02	0.01	[0.01, 0.03]	3.27	.001
Task [narrative]	0.01	0.01	[0.00, 0.02]	2.08	.038
Round [In-sojourn_2]	-0.01	0.01	[-0.02, -0.00]	-2.45	.015
Round [L1]	-0.02	0.01	[-0.03, -0.01]	-3.29	.001
Syntactic complexity	0.01	0.00	[0.01, 0.02]	4.51	<.001
L2Group [Spanish] × Task [narrative]	-0.02	0.01	[-0.04, -0.01]	-3.19	.002
L2Group [Spanish] × Round [In-sojourn_2]	-0.01	0.01	[-0.02, 0.01]	-1.02	.311
L2Group [Spanish] × Round [L1]	-0.02	0.01	[-0.04, -0.01]	-3.01	.003
Task [narrative] × Round [In-sojourn_2]	-0.01	0.01	[-0.03, 0.00]	-1.75	.081
Task [narrative] × Round [L1]	-0.04	0.01	[-0.05, -0.02]	-5.31	<.001
(L2Group [Spanish] × Task [narrative]) × Round [In-sojourn_2]	0.01	0.01	[-0.01, 0.03]	0.70	.483
(L2Group [Spanish] × Task [narrative]) × Round [L1]	0.02	0.01	[-0.00, 0.04]	1.77	.077
Random effects					
σ^2	0.00				
τ_{00}	0.00	Participant			
ICC	0.25				
<i>N</i>	49	Participant			
Observations	294				
Marginal R^2 / Conditional R^2	.479 / .608				

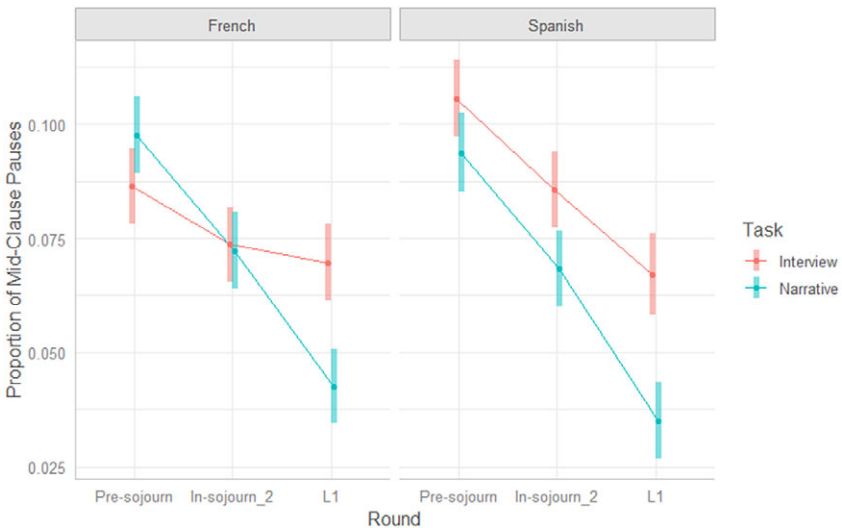


Figure 7. Estimated marginal means and 95% CIs for the final proportion model.

showed improvement on each measure in the narrative task but ultimately remained less fluent in their L2 in comparison with their L1. In the interview task, the only measure of midclause pausing that consistently differentiated L1 from L2 speech was midclause pause rate. Midclause pause rate showed no differences across tasks in the L1.

Discussion

This study set out to investigate the effects of speaking task on midclause pausing characteristics in the L1 and L2 speech of the same speakers to gain further insights into the potential relationship between pause location and stages of speech production. The first research question focused on comparing midclause pausing characteristics in the L1 between a narrative and interview task and considered three types of midclause pause features: the rate (or frequency) of midclause silent pauses, the duration of midclause silent pauses, and the proportion of midclause silent pauses.

The findings indicated that speakers, when using their L1, were more fluent on the narrative task in terms of the duration and proportion of their midclause silent pauses. The difference between tasks was most noticeable in terms of the overall proportion of time spent pausing within a clause. No significant difference was found regarding the frequency of midclause pauses. As argued by Foster and Tavakoli (2009) and Tavakoli and Foster (2008), it is important to have L1 speaker baseline data when attempting to make claims about differences in L1 versus L2 speech-production processes. The fact that the speaking task affected fluency for some midclause pausing measures even when speakers were speaking in their L1 likely supports a more nuanced interpretation of what midclause pauses might represent when considering models of speech production. It has been hypothesized that being less fluent in terms of midclause pausing may be indicative of L2 speech (in comparison with L1 speech) because learners likely do not have as substantial a lexicon and/or efficient access to it as L1 speakers do (Kormos, 2006; Skehan et al., 2016). In other words, midclause pausing has been linked to formulation difficulties. Multiple explanations might account for why speakers in their L1 demonstrate fluency differences between narrative and interview tasks. For instance, pausing for longer stretches within clauses during an interview task, although less likely to represent formulation difficulties for L1 speakers in comparison with L2 speakers, might be connected to the monitor and/or increased reformulation. Recall the importance Segalowitz (2016) placed on attentional/intentional demands in communication. During the interview, the participants were conveying information about their personal opinions and experiences that was unknown to their interlocutors. In contrast, during the narrative task, even though such tasks are designed to put speakers in a position to convey a message, participants were likely aware that their interlocutors were familiar with the stories. Thus, the interview task might invoke stronger demands on speakers to “establish joint attention and [read] each other’s social intentions” (Segalowitz, 2016, p. 88) which in turn might result in increased monitoring and reformulation in light of interlocutors’ verbal and nonverbal feedback. Future research in this area carefully manipulating such task design features could shed more light on these questions.

Another finding from research question one is that not all aspects of midclause pausing showed differences in L1 speech across the two tasks. L1 fluency differences between the two tasks were evident in the measure of proportion and duration but not for the measure of rate. Regarding rate, it may be important to consider that the frequency of midclause pausing in both tasks was relatively low. This finding supports previous work that has indicated that L1 speakers are less likely to pause within clauses (Goldman-Eisler, 1972; Pawley & Syder, 2000). Regarding proportion, which considers the relative amount of time spent pausing within and between clauses, the results indicated that a higher proportion of the total silent pausing occurred at clause boundaries in the narrative task compared with in the interview task. The proportion values were necessarily normalized by clause length, but to put the results in more easily interpretable terms, the raw values indicated that only about one quarter of pausing

(in terms of duration) occurs within clauses in the narrative, whereas this value increased to approximately one half in the interview task. One potential explanation for this difference connects to the discussion in the previous paragraph regarding the interview task invoking stronger attentional/intentional demands and thus increased reformulation within clauses. Presuming a clustering effect of disfluencies, increased reformulations may have come with proportionally longer silent pausing within clauses on the interview. Future work might explore the effects of task on reformulation as a way to begin answering this question. As task differences appeared to affect L1 pausing characteristics most in terms of proportion and least in terms of rate, one practical implication might be that future research on L2 speech incorporates measures of midclause pause rate, as those seem more stable in L1 speech. For instance, it would be interesting to discover whether the midclause pause findings of Kahng (2020) and Suzuki and Kormos (2023) would be even stronger if a measure of proportion was examined (as measures in those studies both focused on rate).

A final consideration regarding research question one relates more practically to design issues that surface when attempting to explore utterance fluency of speakers in their first and second languages longitudinally. For instance, the LANGSNAP corpus collected L1 and L2 speech data at different points using the same narrative task. The same picture narrative was used to avoid potential complications: If a different prompt had been employed, any L1–L2 differences might have occurred because the new task differed based on internal characteristics (e.g., perhaps the vocabulary necessary to complete was more difficult). The choice to employ an existing narrative from the project was done with care: The chosen narrative was selected because it had been the longest time since participants had completed that task—approximately one full year—which allowed for maximal avoidance of any practice effects. Similarly, the L1 data were collected later in the project based on the notion that practice effects would more likely affect the L2 than the L1: Speaker fluency might be more stable in the L1 (in comparison with the L2) because linguistic knowledge and access to it is likely more robust and efficient in the L1 for the current population (adult, instructed L2 learners). That being said, this raises an interesting question regarding the nature of crosslinguistic influence and the potential effects of immersion experiences on midclause pausing characteristics. For instance, a growing body of literature has attested to L2 effects on the L1 at the phonetic level (see e.g., Kartushina et al., 2016) particularly in extensive immersion contexts with potential L1 attrition such as emigrant populations living in the L2 environment for 15+ years (Bergmann et al., 2016). Whether and how more global aspects of speech such as the L1–L2 (dis)fluency characteristics explored in the current study might be similarly affected by immersion experiences is an empirical question warranting future research.

The second research question investigated the extent to which the rate, duration, and proportion of midclause silent pauses in L2 speech changed in a narrative and interview task as proficiency increased and compared these with the same speakers' midclause pausing characteristics in their L1. For the narrative task, all three aspects of midclause pausing improved in the L2 over time; however, the speakers remained less fluent in their L2 in comparison with their L1. For the interview task, although that same trend was found for rate, no differences were evident for duration. For proportion, the results were slightly mixed such that there was some indication of differences between the L2 groups. Although any L1–L2 differences that existed for proportion at *Presojourn* were no longer present by *In-sojourn 2* for the French group, the Spanish group showed some remaining L1–L2 differences (although the CIs were close to overlapping). The fact that speakers' fluency in their L2 improved during residence abroad provides

additional support for the relatively robust finding in the literature that study abroad can positively affect oral production (e.g., Du, 2013; Huensch & Tracy-Ventura, 2017a; Segalowitz & Freed, 2004). Improvement in the L2 was demonstrated in both the narrative and interview task with the largest effects evident for rate.

Regarding the use of midclause pausing as a useful measure of differentiating L1 from L2 speech, on one hand the findings of the current study corroborate previous studies (e.g., De Jong, 2016; Kahng, 2014) who have reported L1–L2 differences. However, L1–L2 differences were not found in all measures in both tasks in the current study. Specifically, only rate remained as a differentiator of L1–L2 speech on the interview task. This means that although speakers were pausing midclause for similar durations overall, the L2 speakers were doing so more frequently. Future research looking to compare L1–L2 speech, thus, might consider using midclause pause rate as a measure, as it consistently differentiated L1 from L2 speech in the current study (and was the only measure to differentiate L1 speech between the two tasks).

Another clear finding from this study is that speaking task had an influence on midclause pausing characteristics: (a) in their L1, participants were more fluent on the narrative task than the interview task and (b) in their L2, participants did not reach L1-like fluency in the narrative task with any midclause pausing measure. It is possible to think about this result in terms of both the narrative task being relatively easier than the interview for speakers in their L1 and/or the narrative task being relatively more difficult than the interview task for the speakers in their L2. Previous research investigating how different elements of narrative tasks affect fluency might offer some potential explanations for the differences that emerged in the current exploration (Derwing et al., 2004; Skehan & Foster, 2012). For instance, the narrative task requires the inclusion of certain lexical items or structures to successfully retell the story, whereas the same is not true for the interview. In this way, the interview task could have resulted in more fluent performance for the speakers in their L2 (similar to that of their L1) because they had more control over what they said and how they said it. Given the number of possible differences between the two tasks employed in the LANGSNAP corpus, future work should carefully manipulate design features to zero in on those with the most influence and with an eye on expanding the scope beyond monologic, narrative tasks.

It is important to acknowledge some potential limitations of the current study. Given the current study's focus on midclause silent pauses, one interesting avenue for future research is to explore whether filled pauses would result in similar findings especially given that previous research has indicated potential cross-language (e.g., De Leeuw, 2007, for English and German L1; Huensch & Tracy-Ventura, 2017b, for French and Spanish L1) and individual differences (Belz et al., 2017) with respect to filled pauses frequency and distribution. Another consideration is related to using existing corpora to answer novel research questions. On one hand, the growing number of rich, publicly available learner corpora is allowing new avenues of research to be explored with existing resources, but they might also have limitations. In using such existing data sets, it is important to acknowledge these limitations and consider approaches to address them, such as the incorporation of lexical and syntactic complexity measures in the current analysis. The findings from the current study provide preliminary indications that midclause silent pausing might be influenced by task effects. This gives support to future work that experimentally manipulates task design features (see Felker et al., 2019, as a nice example) to further tease apart these variables, armed with the findings that different aspects of midclause silent pause (e.g., rate vs. proportion) were not equally affected.

Conclusion

The current study provided a detailed treatment of a single aspect of utterance fluency—midclause silent pausing—and explored the potential effects of speaking task on L1 and L2 fluency. The examination of midclause silent pauses considered frequency, duration, and proportion. In short, speaking task was shown to affect midclause pausing behavior in both L1 and L2 speech, and not all measures of midclause pausing were equally affected. Broadly speaking, these findings have potential implications for second language assessment and pedagogy. For instance, in following Kahng (2018) and De Jong (2016) in arguing for the importance of pause location information as it pertains to language assessment tools, the current study provides preliminary evidence that midclause silent pause rate (or a measure of frequency) might be most appropriate as opposed to a measure of pause duration. More specifically, the findings are relevant for L2 fluency and speech production research as between-task differences in L1 speech call into question what midclause silent pauses might represent in terms of speech production processes.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0272263123000323>.

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at Data: <https://osf.io/dn6v3/>; Materials: <https://www.iris-database.org/details/yhzL0-J2jzK>

Acknowledgments. LANGSNAP was funded by the ESRC (award number RES-062-23-2996). I am grateful to my colleagues Nicole Tracy-Ventura, Kevin McManus, Rosamond Mitchell, and the rest of the LANGSNAP team. I am also thankful to the participants, the transcribers, and my research assistants, in particular Aneesa Ali, for their contribution to this work.

Competing interest. The author declares none.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4* (R package version 1.1.-7) [Computer software]. CRAN.R-project.org/package = lme4
- Belz, M., Sauer, S., Lüdeling, A., & Mooshammer, C. (2017). Fluently disfluent? Pauses and repairs of advanced learners and native speakers of German. *International Journal of Learner Corpus Research*, 3, 118–148. <https://doi.org/10.1075/ijlcr.3.2.02bel>
- Bergmann, C., Nota, A., Sprenger, S. A., & Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics*, 58, 71–86. <https://doi.org/10.1016/j.wocn.2016.07.001>
- Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* (Version 5.4.18) [Computer program]. Retrieved from <http://www.praat.org/>
- Bowden, H. (2016). Assessing second-language oral proficiency for research: The Spanish elicited imitation task. *Studies in Second Language Acquisition*, 38, 647–675. <https://doi.org/10.1017/S0272263115000443>
- Davies, A. (2003). *The native speaker: Myth and reality*. Multilingual Matters.
- De Bot, K. (1992). A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics*, 13, 1–24. <https://doi.org/10.1093/applin/13.1.1>
- De Clercq, B., & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. *Modern Language Journal*, 101, 315–334. <https://doi.org/10.1111/modl.12396>
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54, 113–132. <https://doi.org/10.1515/iral-2016-9993>
- De Jong, N. H., & Bosker, H. R. (2013, August 13–21). *Choosing a threshold for silent pauses to measure second language fluency* [Paper presentation]. The 6th Workshop on Disfluency in Spontaneous Speech, Stockholm, Sweden.

- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36, 223–243. <https://doi.org/10.1017/S0142716413000210>
- De Jong, N. H., & Mora, J. C. (2019). Does having good articulatory skills lead to more fluency speech in first and second languages? *Studies in Second Language Acquisition*, 41, 227–239. <https://doi.org/10.1017/S0272263117000389>
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34, 893–916. <https://doi.org/10.1017/S0142716412000069>
- De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, 19, 85–114. <https://doi.org/10.1017/S1470542707000049>
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533–557. <https://doi.org/10.1017/S0272263109990015>
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54, 655–679. <https://doi.org/10.1111/j.1467-9922.2004.00282.x>
- Domínguez, L., Tracy-Ventura, N., Arche, M., Mitchell, R., & Myles, F. (2013). The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition*, 16, 558–577. <https://doi.org/10.1017/S1366728912000363>
- Du, H. (2013). The development of Chinese fluency during study abroad in China. *Modern Language Journal*, 97, 131–143. <https://doi.org/10.1111/j.1540-4781.2013.01434.x>
- Felker, E., Klockmann, H. E., & De Jong, N. H. (2019). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, 40, 111–136. <https://doi.org/10.1017/S0142716418000474>
- Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*, 58, 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323. <https://doi.org/10.1017/S0272263100015047>
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59, 866–896. <https://doi.org/10.1111/j.1467-9922.2009.00528.x>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. <https://doi.org/10.1093/applin/21.3.354>
- Garrod, S. (1999). The challenge of dialogue for theories of language processing. In S. Garrod & M. J. Pickering (Eds.), *Language processing* (pp. 180–193). Psychology Press.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103–113. <https://doi.org/10.1177/002383097201500201>
- Huensch, A., & Tracy-Ventura, N. (2017a). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, 101, 275–293. <https://doi.org/10.1111/modl.12395>
- Huensch, A., & Tracy-Ventura, N. (2017b). Understanding L2 fluency behavior: The effects of individual differences in L1 fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 38, 755–785. <https://doi.org/10.1017/S0142716416000424>
- Huensch, A., Tracy-Ventura, N., Bridges, J., & Cuesta-Medina, J. (2019). Variables affecting the maintenance of L2 proficiency and fluency four years post-study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 4, 96–125. <https://doi.org/10.1075/sar.17015.hue>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64, 809–854. <https://doi.org/10.1111/lang.12084>
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39, 569–591. <https://doi.org/10.1017/S0142716417000534>
- Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, 41, 457–480. <https://doi.org/10.1017/S0142716420000065>

- Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, 66, 155–186. <https://doi.org/10.1111/lang.12187>
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates Publishers.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Langley, J. (2006). *I very really miss you (illustrations)*. Francis Lincoln.
- Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.
- Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown and P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford University Press.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.
- MacWhinney, B. (2017). A shared platform for studying second language acquisition. *Language Learning*, 67, 254–275. <https://doi.org/10.1111/lang.12220>
- Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 141–173). John Benjamins.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships and language learning*. Routledge.
- Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners* [Unpublished Doctoral dissertation]. University of Hawaii, Honolulu.
- Pawley, A., & Syder, F. (2000). The one clause at a time hypothesis. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 163–191). University of Michigan Press.
- Peltonen, P. (2017). Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System*, 70, 1–13. <https://doi.org/10.1016/j.system.2017.08.009>
- Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *Modern Language Journal*, 102, 676–692. <https://doi.org/10.1111/modl.12516>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R² values. *Modern Language Journal*, 102, 713–731. <https://doi.org/10.1111/modl.12509>
- R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. www.R-project.org/
- Sato, M. (2014). Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System*, 45, 79–91. <https://doi.org/10.1016/j.system.2014.05.004>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54, 79–95. <https://doi.org/10.1515/iral-2016-9991>
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–199. <https://doi.org/10.1017/S0272263104262027>
- Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research-. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & S. C. Vedder (Eds.), *Complexity, accuracy, and fluency in second language use, learning, and teaching* (pp. 207–226). John Benjamins.
- 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. <https://doi.org/10.1515/iral-2016-9992>
- Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 42, 143–167. <https://doi.org/10.1017/S0272263119000421>
- Suzuki, S., & Kormos, J. (2023). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, 45, 38–64. <https://doi.org/10.1017/S0272263121000899>
- Tavakoli, P. (2011). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, 65, 71–79. <https://doi.org/10.1093/elt/ccq020>

- Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 133–150. <https://doi.org/10.1515/iral-2016-9994>
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58, 439–473. <https://doi.org/10.1111/j.1467-9922.2008.00446.x>
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). John Benjamins.
- Tracy-Ventura, N., & Huensch, A. (2018). The potential of publicly shared longitudinal learner corpora in SLA research. In A. Gudmestad & A. Edmonds (Eds.), *Critical Reflections on data in second language acquisition* (pp. 149–170). John Benjamins.
- Van Os, M., De Jong, N. H., & Bosker, H. R. (2020). Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and nonnative speech. *Language Learning*, 70, 1183–1217. <https://doi.org/10.1111/lang.12416>

Appendix: Semistructured interview questions

Presojourn

1. Pourquoi as-tu choisi d'étudier les langues vivantes ? / ¿Por qué decidiste estudiar idiomas?
Why did you decide to study foreign languages?
2. Quels buts as-tu pour toi-même pendant l'année à l'étranger ? Développement linguistique/culturel/personnel, indépendance? / ¿Tienes algún proyecto u objetivo personal que quieras lograr durante tu año en el extranjero? ¿Ya sea lingüístico, cultural, personal, del vivir de manera independiente, etc.?
Do you have any personal objectives or goals you want to achieve during your year abroad? Linguistic, cultural, or personal development, learning to be independent?

In-sojourn 2

1. Qu'est-ce qui s'est passé depuis ta dernière visite en novembre ? / Cuéntame ¿qué ha pasado desde mi última visita/la última visita que te hice?
What has happened since your last visit?
2. Est-ce que tu penses que ton niveau de français s'est amélioré depuis ta dernière visite en novembre ? / ¿Crees que tu español haya mejorado desde mi última visita?
Do you think your French/Spanish has improved since the last visit?

English (L1)

1. Now that you are at the end of your year abroad, how do you think the experience has influenced your learning of French/Spanish ?
2. What kinds of frustrations did you encounter learning the language in France/Spain/Mexico (if any)?

Cite this article: Huensch, A. (2023). Effects of speaking task and proficiency on the midclause pausing characteristics of L1 and L2 speech from the same speakers. *Studies in Second Language Acquisition*, 45: 1031–1055. <https://doi.org/10.1017/S0272263123000323>